NPSOR-93-003

NAVAL POSTGRADUATE SCHOOL

Monterey, California

PRELIMINARY RESULTS FROM THE ANALYSIS OF WIND COMPONENT ERROR JULY DATA

Donald P. Gaver
Patricia A. Jacobs

November 1992

Approved for public release; distribution is unlimited.

Prepared for:
Naval Research Laboratory-West
›nterey, CA 93943-5006

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral R. W. West, Jr.
Superintendent

Harrison Shull
Provost

This report was prepared in conjunction with research funded by the Naval Research Laboratory-West, Monterey, California.

This report was prepared by:

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution is unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
NPSOR-93-003

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
Naval Postgraduate School

6b. OFFICE SYMBOL (If applicable)
OR

7a. NAME OF MONITORING ORGANIZATION
Naval Research Laboratory-West

6c. ADDRESS (City, State, and ZIP Code)
Monterey, CA 93943

7b. ADDRESS (City, State, and ZIP Code)
Monterey, CA 93943-5006

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
Naval Research Laboratory-West

8b. OFFICE SYMBOL (If applicable)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (City, State, and ZIP Code)
Monterey, CA 93943-5006

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
Preliminary Results from the Analysis of Wind Component Error-July Data

12. PERSONAL AUTHOR(S)
Donald P. Gaver and Patricia A. Jacobs

13a. TYPE OF REPORT
Technical Report

13b. TIME COVERED
FROM______TO______

14. DATE OF REPORT (Year, month day)
November 1992

15. PAGE COUNT
107

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Gaussian model with log-linear scale parameter; maximum likelihood; Newton's method

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Estimation of mean square prediction error of wind components is required in the optimal interpolation (OI) process in numerical prediction of atmospheric variables. Statistical models with log-linear scale parameters which include covariates are described for the prediction error. Data from February April and July of 1991 are used to fit the model parameters and to study the predictive ability of the models. This preliminary investigation indicates that observational and first guess wind components can be helpful in predicting mean square prediction error for wind components. The predictions using observational winds appear to be better at the 850 mb level. The predictions using first guess winds appear to be better at the 250 mb level.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
[x] UNCLASSIFIED/UNLIMITED [ ] SAME AS RPT. [ ] DTIC USERS

21. ABSTRACT SECURITY CLASSICIATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
D. P. Gaver

22b. TELEPHONE (Include Area Code)
(408) 646-2605

2c. OFFICE SYMBOL
OR/Gv

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603 UNCLASSIFIED

PRELIMINARY RESULTS FROM THE ANALYSIS OF
WIND COMPONENT ERROR
JULY DATA

by

P. A. Jacobs and D. P. Gaver

TABLE OF CONTENTS

PRELIMINARY RESULTS FROM THE ANALYSIS OF WIND COMPONENT ERROR JULY DATA

by

P. A. Jacobs and D. P. Gaver

0. INTRODUCTION

Numerical meteorological models are used to assist in the prediction of weather. Each run of a numerical model produces forecasts of meteorological variables which are used as preliminary predictions of the future values of these variables. These initial predictions are referred to as **first-guess** values. In this paper first-guess values will refer to the most recent 12 hour forecasts.

In certain areas of the world, observations of the values of forecasted variables become available. In our case the observations become available 12 hours after the first-guess values are computed. Prior to the next run of the numerical model a multivariate optimal interpolation analysis updates a first-guess value of a variable by adding to it a weighted observed value of the variable if it is available. The weight multiplying the observed value depends on estimates of the squared error of the first-guess value and the squared error of the observation; cf. Goerss et al. [1991, a, b]. Thus it is of importance to predict such first-guess squared errors.

The general problem of modeling and predicting mean square errors is important but not widely studied; see Davidian and Carroll (1987), Nelder and Lee (1992), Aitken (1987), and McCullagh and Nelder (1983). In the next section statistical models for the error of the first-guess are introduced. The models assume the error of the first-guess has mean 0 but has a scale parameter that is log-linear with suitable covariates, i.e. explanatory or regression variables.

Results are reported concerning the estimation of model parameters, and model cross-validation and predictive ability for u, v wind component data from the month of July 1991. The data consist of measurements and 12 hour forecasts (first-guess values) at the 850 mb, 500 mb and 250 mb levels from 93 stations in North America, 25N–75N. The forecasts are produced using the NOGAPS Spectral Forecast Model; cf. Hogan et al. (1991). Each station has measurement and first-guess values for every 12 hours; there are some missing observations and suspicious values of wind components equal to 0. These missing and questionable values are deleted from the data set. The measurement values (if available) are subtracted from first-guess values to obtain observations of the error of the first-guess value. The results appear in Sections 3 and 4 and in Appendices A, B.

The results indicate that estimates of the variance of the error of first-guess wind components can be improved by using covariates which are functions of the wind components. Covariates using observed values of the wind components appear to have more predictive ability than those using first-guess values. Further exploratory work is needed to determine the degree with which these statistical results can be used to improve the forecasting ability of the numerical model.

1. THE MODELS

Fix a location. Let

$U_0(t)$ = observed u-wind component at time t

$U_f(t)$ = first-guess u-wind component at time t

$V_0(t)$ = observed v-wind component at time t

$V_f(t)$ = first-guess v-wind component at time t

$$r(t) = \left[\left(U_0(t) - U_0(t-1) \right)^2 + \left(V_0(t) - V_0(t-1) \right)^2 \right]^{\frac{1}{2}}$$

$$s(t) = \left[U_0(t)^2 + V_0(t)^2 \right]^{\frac{1}{2}}$$

$$Y(t) = U_0(t) - U_f(t) \quad \text{or} \quad Y(t) = V_0(t) - V_f(t)$$

The variable $Y(t)$ is the first guess error. The variable $r(t)$ is a measure of the observed change in the wind. The variable $s(t)$ is the observed wind speed.

The models considered are as follows:

One Variable Models

1. $\{Y(t)\}$ are independent normally distributed random variables with mean 0 and variance

$$\sigma_1^2(1;t) = \exp\{\alpha_1(1) + \beta_1(1) r(t)\}. \tag{1}$$

2. $\{Y(t)\}$ are independent normally distributed random variables with mean 0 and variance

$$\sigma_1^2(2;t) = \exp\{\alpha_1(2) + \beta_1(2) s(t)\}. \tag{2}$$

Two Variable Model

3. $\{Y(t)\}$ are independent normally distributed random variables with mean 0 and variance

$$\sigma_2^2(t) = \exp\{\alpha + \beta_1 r(t) + \beta_2 s(t)\}. \tag{3}$$

Independence Assumption

The first guess errors at different locations are independent. The parameters in the variance models do not depend on location.

2. ESTIMATION OF PARAMETERS

The model parameters are estimated by maximum likelihood. A system of equations is obtained by setting the first partial derivative with respect to each parameter of the log likelihood function equal to zero. The system of equations is solved numerically using Newton's method to obtain the maximum likelihood estimates. The procedure for the normal models above is given in Appendix A of Jacobs and Gaver [1991].

3. THE DATA ANALYSIS—JULY DATA

3.1 Observed Wind Covariate Models

In this subsection we report an assessment of the goodness of fit and cross-validation for the normal models (1)–(3) using observational wind components as covariates. There are six analyses; one for the u-wind component (respectively v-wind component) for each pressure level. Once missing values and suspicious wind values of 0 are deleted there are 3519 data values at the 850 mb level, 3833 values at the 500 mb level and 3830 values at the 250 mb level. Each analysis proceeds along the same lines. In what follows by data we mean triples $\{y(t), r(t), s(t)\}$.

In each analysis the data are randomly divided into two sets called DA and DB without regard to the values of the data.

The maximum likelihood parameter estimates for each model (1)–(3) are obtained for each set DA and DB and for all the data. The parameter estimates and their estimated standard errors (computed from the second partial derivatives of the likelihood evaluated at the estimates) appear in Table 1. Note that all the estimates are positive. Hence increased $r(t)$ and/or $s(t)$ values are associated with increased variance of the first guess value. This is plausible physically, since a large value of $r(t)$ is indicative of a change in the atmosphere and a large value of $s(t)$ is indicative of an active location in the atmosphere. The estimated variances $\sigma_1^2(1,t)$, $\sigma_1^2(2,t)$, $\sigma_2^2(t)$, are computed for the parameters estimated from DA and DB using (1)–(3) for each data point in DA and DB.

The models are for the variances of the observations rather than the observations themselves. One possible procedure to informally assess goodness-of-fit and cross-validate the models is by binning the data. To assess models (1) and (3) the data $(y(t), r(t), s(t))$ are binned into 10 bins based on ordering the values of $r(t)$ from smallest to largest. The data in the first bin correspond to the smallest values of $r(t)$; the data in the 10th bin correspond to the largest values of $r(t)$. Each bin contains about $\frac{1}{10}$th of the data with the 10th bin containing a few more data. The averages of the estimated variances for models (1) and (3) are computed for each bin. The average $y(t)^2$ is also computed for each bin.

To assess models (2) and (3) the same procedure is used but the binning is based on the values of $s(t)$.

TABLE 1. NORMAL MODELS
JULY DATA PARAMETER ESTIMATES
(STANDARD ERRORS)
OBSERVED WIND COVARIATES

			One-variate Models				Two-variate Models		
Pressure Level	Wind Comp.	Data Set	$r(t)$ α	$r(t)$ β	$s(t)$ α	$s(t)$ β	log MSE = $\alpha+\beta_1 r(t)+\beta_2 s(t)$ α	β_1	β_2
850	u	A	1.47 (0.06)	0.11 (0.008)	1.50 (0.06)	0.09 (0.007)	1.25 (0.06)	0.09 (0.01)	0.05 (0.009)
		B	1.42 (0.06)	0.12 (0.010)	1.41 (0.06)	0.10 (0.008)	1.13 (0.07)	0.08 (0.01)	0.07 (0.009)
		ALL	1.45 (0.04)	0.11 (0.006)	1.46 (0.04)	0.09 (0.005)	1.20 (0.05)	0.08 (0.007)	0.06 (0.006)
850	v	A	1.58 (0.06)	0.10 (0.008)	1.53 (0.06)	0.09 (0.008)	1.36 (0.07)	0.07 (0.01)	0.05 (0.009)
		B	1.51 (0.06)	0.11 (0.009)	1.52 (0.06)	0.10 (0.008)	1.21 (0.007)	0.09 (0.01)	0.06 (0.009)
		ALL	1.55 (0.04)	0.11 (0.006)	1.53 (0.04)	0.09 (0.006)	1.29 (0.05)	0.08 (0.007)	0.06 (0.006)
500	u	A	1.37 (0.06)	0.12 (0.008)	1.54 (0.06)	0.06 (0.005)	1.12 (0.07)	0.10 (0.008)	0.03 (0.005)
		B	1.45 (0.06)	0.10 (0.009)	1.66 (0.06)	0.04 (0.005)	1.24 (0.07)	0.09 (0.009)	0.02 (0.005)
		ALL	1.40 (0.04)	0.11 (0.006)	1.58 (0.04)	0.05 (0.005)	1.17 (0.05)	0.10 (0.006)	0.03 (0.004)
500	v	A	1.53 (0.06)	0.09 (0.009)	1.74 (0.06)	0.03 (0.005)	1.35 (0.08)	0.08 (0.009)	0.02 (0.005)
		B	1.45 (0.06)	0.11 (0.009)	1.59 (0.06)	0.05 (0.005)	1.20 (0.07)	0.09 (0.009)	0.03 (0.005)
		ALL	1.49 (0.04)	0.10 (0.006)	1.66 (0.05)	0.04 (0.004)	1.27 (0.05)	0.09 (0.007)	0.03 (0.004)
250	u	A	2.41 (0.06)	0.06 (0.005)	2.50 (0.06)	0.03 (0.003)	2.13 (0.07)	0.05 (0.005)	0.02 (0.003)
		B	2.42 (0.06)	0.06 (0.005)	2.53 (0.06)	0.03 (0.003)	2.13 (0.07)	0.05 (0.005)	0.02 (0.003)
		ALL	2.41 (0.04)	0.06 (0.003)	2.52 (0.04)	0.03 (0.002)	2.13 (0.05)	0.05 (0.004)	0.02 (0.002)
250	v	A	2.41 (0.06)	0.07 (0.005)	2.47 (0.06)	0.03 (0.003)	2.08 (0.07)	0.05 (0.005)	0.02 (0.003)
		B	2.49 (0.06)	0.05 (0.005)	2.39 (0.06)	0.03 (0.003)	2.16 (0.07)	0.04 (0.005)	0.02 (0.003)
		ALL	2.44 (0.04)	0.06 (0.003)	2.43 (0.04)	0.03 (0.002)	2.12 (0.05)	0.05 (0.003)	0.02 (0.002)

$r(t) = [((u(t) - u(t-1))^2 + (v(t) - v(t-1))^2)]^{1/2}$

$s(t) = [u(t)^2 + v(t)^2]^{1/2}$

NOTE: Data are divided into two sets randomly without regard to data values. One set is called A; the other is called B.

Figures 1-24 present graphs of the log [average $y(t)^2$] in each bin versus log [average estimated variance] in each bin for models (1) and (3) and models (2) and (3). Figures 1, 5, 9, 13, 17, 21 (respectively 2, 6, 10, 14, 18, 22) show the logarithm of the average of the $y(t)^2$ values of DA (respectively DB) versus the logarithm of the average of the estimated variances for each bin using the estimated parameters from DA (respectively DB). If a model were perfect, a point should be close to the 45° line shown.

Figures 3, 7, 11, 15, 19, 23, (respectively 4, 8, 12, 16, 20, 24) present graphs of log average $y(t)^2$ of DA (respectively DB) versus log average estimated variances using parameters estimated using data DB (respectively DA). Once again if the model were perfect, the points would be close to the 45° line.

Since the two-variate model (3) is shown with both one-variate models, it is possible to obtain some idea of the effect of the two different sets of bins on the log averages. In particular, the graphs corresponding to the 500 Mb height winds, Figures 9-16, show that the display of log averages can be quite sensitive to which variate is used to do the binning.

Keeping this binning sensitivity in mind, the figures suggest the following concerning the models using observed winds as covariates. It appears that of the two one-variate models, model (1) which uses $r(t)$ as the covariate is the better. The two-variate model (3) appears somewhat better than model (1). If wind speed is used as the single covariate, it appears to overstate the variance; the addition of the second covariate $r(t)$ in this case seems to tend to make the estimated variance smaller and bring the log [average predicted variance] in a bin closer to the log average y^2 in the bin.

Another way to assess goodness of fit and to cross validate is to evaluate the log-likelihood for the different models at the parameter estimates. Larger values of the log-likelihood suggest better model fit; cf. Cox and Hinkley [1974].

Table 2 presents the values of the log-likelihood at the parameter estimates up to addition and multiplication of constants for the parameter estimates of Table 1; the function being evaluated is

$$\tilde{\ell} = -n\alpha - \sum_{i=1}^{n} \underline{x}_i \underline{\beta} - \sum_{i=1}^{n} y_i^2 \exp\left\{-\alpha + \underline{x}_i \underline{\beta}\right\}. \tag{4}$$

where $\underline{x}_i \underline{\beta} = \sum_j x_{ij}\beta_j$. The values of $\tilde{\ell}$ are presented for data DA (respectively DB) using the parameters fit using DA (respectively DB); these are values assessing goodness of fit; since maximum likelihood is the estimation procedure, the largest value of $\tilde{\ell}$ in each of these two rows is the one corresponding to the two-variate model. Values of $\tilde{\ell}$ are also presented for data DA (respectively DB) using the parameters fit using DB (respectively DA); these are values assessing cross-validation. The underlined value in each row is the maximum value in that row; the corresponding model provides the best model fit. The bold italicized value in each row is the maximum value for the two one-variate models; the corresponding one-variate model provides the best model fit between the two one-variate models.

The models considered in Table 2 are models (1)-(3) and the model that the $\{Y_i\}$ are independent normal with mean 0 and variance not a function of the covariates; that is,

TABLE 2. NORMAL MODELS
JULY DATA
OBSERVED WIND COVARIATES
LOG-LIKELIHOOD

Pressure Level	Wind Comp.	Data Set	Model	Constant	One-variate Models $r(t)$	$s(t)$	Two-variate Models
850	u	A	A	–5727.0	*–5424.4*	–5468.8	–5386.5
		B	B	–5337.4	*–5362.1*	–5379.5	–5306.5
		B	A	–5546.9	*–5363.6*	–5381.8	–5310.7
		A	B	–5737.2	*–5425.9*	–5471.5	–5391.2
850	v	A	A	–5680.1	*–5486.5*	–5500.9	–5448.5
		B	B	–5693.0	*–5504.0*	–5540.3	–5450.2
		B	A	–5693.1	*–5506.9*	–5542.6	–5457.8
		A	B	–5680.1	*–5489.7*	–5503.4	–5456.6
500	u	A	A	–6237.3	*–5909.8*	–6049.7	–5871.5
		B	B	–5958.0	*–5821.1*	–5892.6	–5795.8
		B	A	–5977.0	*–5827.4*	–5912.7	–5802.1
		A	B	–6258.2	*–5918.7*	–6076.9	–5879.9
500	v	A	A	–6023.9	*–5904.1*	–5976.2	–5889.2
		B	B	–6193.6	*–5997.8*	–6072.8	–5961.6
		B	A	–6201.7	*–6005.0*	–6090.1	–5971.2
		A	B	–6031.5	*–5910.2*	–5990.6	–5897.3
250	u	A	A	–7893.9	*–7680.0*	–7762.9	–7631.9
		B	B	–7981.7	*–7760.4*	–7829.5	–7703.2
		B	A	–7983.7	*–7760.7*	–7830.2	–7703.4
		A	B	–7895.9	*–7680.3*	–7763.6	–7632.1
250	v	A	A	–8025.3	*–7770.6*	–7850.5	–7710.6
		B	B	–7758.0	*–7611.7*	–7622.3	–7554.3
		B	A	–7775.8	*–7624.4*	–7639.0	–7569.5
		A	B	–8044.9	*–7786.3*	–7868.7	–7729.5

$$\sigma_i^2(t) = e^{\alpha} \qquad \text{(Constant variance).} \tag{5}$$

The two-variate model (3) maximizes the cross-validation values of $\tilde{\ell}$ for data DA (respectively DB) with a model using parameters fit using DB (respectively DA). This suggests that both $r(t)$ and $s(t)$ together have predictive ability.

For the one-variate models (1) and (2) the cross-validation values of $\tilde{\ell}$ for DA (respectively DB) using the parameters fit using DB (respectively DA) are maximized when $r(t)$ is the variable for all cases. This suggests that $r(t)$ by itself has better predictive value than the wind speed $s(t)$ by itself. The goodness of fit values of $\tilde{\ell}$ for the one-variate models using DA (respectively DB) have a higher value of $\tilde{\ell}$ associated with $r(t)$ the majority of the time. This suggests that $r(t)$ by itself provides a better description of the data than $s(t)$ by itself.

Comparing the value of $\tilde{\ell}$ for the model with constant variance (5), $\tilde{\ell}_c$, for DA (respectively DB) fit using DA (respectively DB) with the corresponding cross-validation value of $\tilde{\ell}$ for DA (respectively DB) using models (2), (3) fit using DB (respectively DA) indicates the following. The values of $\tilde{\ell}$ for models (1), (2) and (3) fit with the other half of the data are larger than the corresponding value $\tilde{\ell}_c$ for the constant variance model fit using the data to be modeled. This indicates that both models (2) and (3) fit with the other half of the data describe the data better than the best constant variance model (5) fit with the same data it is used to summarize.

Table 3 presents values of the fraction of increase in $\tilde{\ell}$, $(\tilde{\ell}-\tilde{\ell}_c)/|\tilde{\ell}_c|$, where $\tilde{\ell}_c$ is the maximum value of $\tilde{\ell}$ for the constant variance model (5) fit using data DA (respectively DB) compared to the cross-validation value of $\tilde{\ell}$ for DA (respectively DB) using models (1)-(3) fit using DB (respectively DA). Large values of the fraction will indicate better model predictive ability. Note that the fraction of increase tends to become larger for higher pressure levels. This behavior suggests that if winds from one pressure level are to be used to estimate the variance of the first guess, it should be the 850 mb level. Comparison of the values for the two one-variate models once again suggests

TABLE 3. JULY OBSERVED WIND COVARIATES
FRACTION OF INCREASE $(\tilde{\ell}-\tilde{\ell}_c)/|\tilde{\ell}_c|$

Pressure Level	Wind Comp.	Data Set	Model	One-variate Models $r(t)$	One-variate Models $s(t)$	Two-variate Models
850	u	B	A	0.03	0.03	0.04
		A	B	0.05	0.04	0.06
	v	B	A	0.04	0.03	0.04
		A	B	0.03	0.03	0.04
500	u	B	A	0.02	0.008	0.03
		A	B	0.05	0.03	0.06
	v	B	A	0.03	0.02	0.04
		A	B	0.02	0.01	0.02
250	u	B	A	0.03	0.02	0.03
		A	B	0.03	0.02	0.03
	v	B	A	0.02	0.02	0.02
		A	B	0.03	0.02	0.04

that the one-variate model using $r(t)$ has the greater predictive ability. Once again the two-variate model appears to have the most predictive ability.

To further explore the predictive ability of the models using observed covariates, bootstrap experiments were conducted. Six bootstrap experiments were conducted; one for each u and v wind component at each pressure level.

Each experiment consists of 250 replications. For each replication the data are randomly divided into two sets independent of their values which we will call A and B. Models (1)-(3) are fit to each data set. The value of $\tilde{\ell}_c$, the value of $\tilde{\ell}$ for the constant variance model fit using the same data it is to describe, is computed for each data set. The value of $\tilde{\ell}$ for each data set is computed for each model (1)–(3) with parameters estimated using the other half of the data. The fraction of increase in $\tilde{\ell}$, $(\tilde{\ell}-\tilde{\ell}_c)/|\tilde{\ell}_c|$ is computed for each half of the data. Figures 1A-6A in Appendix A display histograms of $(\tilde{\ell}-\tilde{\ell}_c)/|\tilde{\ell}_c|$ for models using observed wind covariates. Each histogram includes the fractions for both A and B data sets. The histogram indicated that the models for the 850 mb level have the most predictive ability. Model (3) using both covariates appears to have somewhat better predictive ability. Of the two one-variate models model (1) using $r(t)$ as the covariate clearly has the better predictive ability.

3.2 First-guess Wind Covariate Models

In this section we report the results of using models (1)–(3) and (5) with first-guess winds as covariates; the two covariates considered are

$$r_f(t) = \left[\left(U_f(t) - U_f(t-1)\right)^2 + \left(V_f(t) - V_f(t-1)\right)^2\right]^{\frac{1}{2}}$$

and

$$s_f(t) = \left[U_f(t)^2 + V_f(t)^2\right]^{\frac{1}{2}}.$$

The first guess resultant wind $r_f(t)$ is a measure of forecasted change in the winds. The first guess wind speed $s_f(t)$ is a measure of forecasted activity in the atmosphere. Since observed winds are not available over a great portion of the earth, it is important to have models for predicting the variance of the first-guess values which involve the first-guess values which are always available.

Once missing values and suspicious 0 wind values are deleted, there are 3710 observations at the 850 mb level, 4208 observations at the 500 mb level, and 4132 observations at the 250 mb level. The analysis is the same as in the previous subsection. The data sets DA and DB are the same as those in the previous subsection in each case. The values of the parameter estimates with estimated standard errors appear in Table 4. Note that the estimates are all positive. Hence increased $r_f(t)$ and/or $s_f(t)$ is associated with higher variance of the first guess error. The corresponding values of $\tilde{\ell}$ appear in Table 5. Once again the underlined value of $\tilde{\ell}$ is the largest value in each row; the bold italicized value $\tilde{\ell}$ is the largest value between the two one-variate models.

In all cases the values of $\tilde{\ell}$ for the observed wind covariates are larger than those for the first-guess wind covariates. This suggests that the observed wind components have better predictive and descriptive value than the first-guess wind components.

TABLE 4. NORMAL MODELS PARAMETER ESTIMATES (STANDARD ERROR) FIRST-GUESS WIND COVARIATES

Pressure Level	Wind Comp.	Data Set	One-variate Models $r_f(t)$ α	$r_f(t)$ β	$s_f(t)$ α	$s_f(t)$ β	Two Variate Models log MSE=$\alpha+\beta_1 r_f(t)+\beta_2 s_f(t)$ α	β_1	β_2
850	u	A	2.04 (0.06)	0.05 (0.01)	2.13 (0.06)	0.02 (0.01)	2.01 (0.07)	0.05 (0.01)	0.005 (0.01)
		B	1.96 (0.06)	0.05 (0.01)	2.00 (0.06)	0.02 (0.009)	1.90 (0.07)	0.04 (0.01)	0.01 (0.009)
		ALL	2.00 (0.04)	0.05 (0.009)	2.07 (0.04)	0.02 (0.006)	1.96 (0.05)	0.05 (0.009)	0.009 (0.007)
850	v	A	2.03 (0.06)	0.04 (0.01)	1.96 (0.06)	0.03 (0.009)	1.91 (0.07)	0.02 (0.01)	0.03 (0.009)
		B	1.94 (0.06)	0.07 (0.01)	1.93 (0.06)	0.05 (0.008)	1.82 (0.07)	0.06 (0.01)	0.03 (0.009)
		ALL	1.98 (0.04)	0.06 (0.008)	1.94 (0.04)	0.04 (0.006)	1.86 (0.05)	0.04 (0.009)	0.03 (0.007)
500	u	A	1.87 (0.05)	0.06 (0.01)	1.75 (0.06)	0.03 (0.005)	1.63 (0.07)	0.04 (0.01)	0.03 (0.005)
		B	1.99 (0.06)	0.04 (0.01)	1.89 (0.06)	0.03 (0.005)	1.82 (0.07)	0.02 (0.01)	0.02 (0.005)
		ALL	1.93 (0.04)	0.05 (0.007)	1.82 (0.04)	0.03 (0.003)	1.72 (0.05)	0.03 (0.008)	0.03 (0.004)
500	v	A	1.95 (0.05)	0.05 (0.01)	1.84 (0.06)	0.03 (0.005)	1.75 (0.07)	0.03 (0.01)	0.02 (0.005)
		B	1.99 (0.05)	0.04 (0.01)	2.01 (0.06)	0.01 (0.005)	1.91 (0.07)	0.04 (0.01)	0.009 (0.005)
		ALL	1.97 (0.04)	0.04 (0.007)	1.92 (0.04)	0.02 (0.004)	1.83 (0.05)	0.03 (0.007)	0.02 (0.004)
250	u	A	2.79 (0.06)	0.04 (0.006)	2.64 (0.06)	0.02 (0.003)	2.54 (0.06)	0.02 (0.007)	0.02 (0.003)
		B	3.00 (0.05)	0.03 (0.006)	2.92 (0.06)	0.02 (0.003)	2.84 (0.06)	0.02 (0.005)	0.01 (0.003)
		ALL	3.00 (0.04)	0.03 (0.004)	2.79 (0.04)	0.02 (0.002)	2.70 (0.05)	0.02 (0.006)	0.02 (0.002)
250	v	A	2.81 (0.05)	0.04 (0.006)	2.71 (0.06)	0.02 (0.003)	2.58 (0.07)	0.03 (0.006)	0.02 (0.003)
		B	2.77 (0.05)	0.05 (0.006)	2.70 (0.06)	0.02 (0.003)	2.51 (0.07)	0.04 (0.006)	0.02 (0.003)
		ALL	2.79 (0.04)	0.04 (0.004)	2.71 (0.04)	0.02 (0.002)	2.55 (0.05)	0.03 (0.004)	0.02 (0.002)

TABLE 5.
JULY LOG-LIKELIHOOD
FIRST-GUESS COVARIATES

Pressure Level	Wind Comp.	Data Set	Model	Constant	One-variate Models $r_f(t)$	One-variate Models $s_f(t)$	Two-variate Models
850	u	A	A	–6018.5	***–6000.0***	–6014.4	–5999.7
		B	B	–5857.1	***–5840.3***	–5848.5	–5837.9
		B	A	–5863.9	***–5847.4***	–5856.9	–5846.3
		A	B	–6025.7	*–6007.5*	–6023.6	–6009.0
850	v	A	A	–5900.1	–5890.2	***–5884.7***	–5881.0
		B	B	–6023.1	***–5977.5***	–5987.7	–5967.3
		B	A	–6027.2	***–5990.4***	–5994.1	–5978.4
		A	B	–5904.1	***–5900.2***	–5900.5	–5889.8
500	u	A	A	–6624.7	–6584.1	***–6567.8***	–6550.7
		B	B	–6683.1	–6669.1	***–6653.3***	–6649.0
		B	A	–6683.9	–6674.1	***–6658.8***	–6656.9
		A	B	–6625.5	–6589.5	***–6573.2***	–6559.0
500	v	A	A	–6658.3	–6636.2	***–6622.8***	–6613.9
		B	B	–6656.4	***–6638.1***	–6648.1	–6635.4
		B	A	–6656.4	*–6638.6*	–6655.4	–6643.6
		A	B	–6658.3	–6636.7	***–6631.0***	–6623.1
250	u	A	A	–8484.0	–8437.4	***–8407.3***	–8392.6
		B	B	–8693.2	–8670.7	***–8660.6***	–8652.7
		B	A	–8704.2	–8689.5	*–8680.6*	–8682.5
		A	B	–8494.3	–8454.1	***–8431.0***	–8418.0
250	v	A	A	–8453.7	–8411.9	***–8399.7***	–8380.1
		B	B	–8550.1	***–8486.4***	–8488.8	–8451.2
		B	A	–8552.4	–8491.0	***–8490.1***	–8455.4
		A	B	–8455.9	–8416.5	***–8400.9***	–8384.2

Table 5 also indicates the following results concerning models using first-guess wind covariates. Between the two one-variate models (1) and (2) the one-variate model using first-guess wind speed has the greater $\tilde{\ell}$-value the majority of the time. This suggests that first-guess wind speed alone has somewhat better predictive and descriptive value than $r_f(t)$ alone. The cross-validation values of $\tilde{\ell}$ for data DA (respectively DB) using parameters fit with DB (respectively DA) are maximized in all cases except two for the two-variate model. This suggests that the two-variate model has better predictive ability.

Comparing the values of $\tilde{\ell}$, $\tilde{\ell}_c$, for DA (respectively DB) using the constant variance model (5) fit using DA (respectively DB) with the cross-validation value of $\tilde{\ell}$ for DA (respectively DB) using models (2), (3) fit using DB (respectively DA) indicates the following. The values of $\tilde{\ell}$ for models (1), (2) and (3) fit with the other half of the data are larger in all but two cases than the corresponding value $\tilde{\ell}_c$ for the constant variance model fit using the data to be modeled. This suggests that models (1)–(3) fit with the other half of the data describe the data somewhat better than the best constant variance model (5) fit with the data to be described.

Table 6 presents the fraction of increase in log-likelihood obtained by using models (1)–(3) fit using data DA (respectively DB) to describe data DB (respectively DA) compared to the value of the likelihood obtained by fitting the constant variance model (5) using data DB (respectively DA); Table 6 shows values of $\frac{\tilde{\ell}-\tilde{\ell}_c}{|\tilde{\ell}_c|}$. The results suggest that models using first guess wind have better predictive ability for lower pressure levels. Hence, if only one pressure level is to be used it is suggested that models for either the 500 mb level or 250 mb level be considered.

TABLE 6
JULY DATA
FRACTION OF INCREASE IN LOG-LIKELIHOOD
FIRST GUESS COVARIATES

$$\frac{\tilde{\ell}-\tilde{\ell}_c}{|\tilde{\ell}_c|}$$

Pressure Level	Wind Comp.	Data Set	Model	One-variate Models $r(t)$	$s(t)$	Two-variate Models
850	u	B	A	0.002	0	0.002
		A	B	0.002	*	0.002
	v	B	A	0.005	0.005	0.007
		A	B	*	0.002	0.002
500	u	B	A	0.001	0.004	0.004
		A	B	0.005	0.008	0.010
	v	B	A	0.003	0.000	0.002
		A	B	0.003	0.004	0.005
250	u	B	A	0.000	0.000	0.001
		A	B	0.004	0.006	0.008
	v	B	A	0.007	0.007	0.011
		A	B	0.004	0.006	0.008

$^*\tilde{\ell}_c > \tilde{\ell}$.

To further explore the predictive ability of the models using first guess covariates, bootstrap experiments were conducted. Six experiments were conducted, one for each u and v wind component at each pressure level. Each experiment consists of 250 replications. For each replication the data are randomly divided into two sets, independent of their values, which we will call A and B. Models (1)-(3) are fit to each data set. The value of $\tilde{\ell}_C$, the value of $\tilde{\ell}$ for the constant variance model fit using the same data it is to describe, is computed for each data set. The value of $\tilde{\ell}$ for each data set is computed for each model (1)-(3) fit using the other half of the data. The fraction of increase in $\tilde{\ell}$, $(\tilde{\ell}-\tilde{\ell}_C)/|\tilde{\ell}_C|$, is computed for each half of the data. Figures 7A-12A in

Appendix A display histograms of $(\tilde{\ell}-\tilde{\ell}_c)/|\tilde{\ell}_c|$ for models using first guess wind covariates. Each histogram includes the fractions for both A and B data sets. The histograms indicate that the models using first guess wind covariates do not have as much predictive ability as those using observed wind covariates. The first guess wind covariate models appear to have the most predictive ability at the 500 and 250 mb levels with the models at the 250 mb level being somewhat better. Model (3) using both first guess covariates appears to have the best predictive ability. Of the two one-variate models, Model (1) using $r_f(t)$ as the covariate has the better predictive ability. The predictive ability of the one-variate model using $s_f(t)$ is the most variable.

In summary, based on values of $\tilde{\ell}$, when first-guess winds are used as covariates it appears that the two-variate model using first-guess wind speed at the 250 mb level is an attractive choice for predictive purposes. When observational winds are used as covariates, the two-variate model at the 850 mb level appears to have the best predictive value.

4. A COMPARISON OF MODELS FOR THE MONTHS OF FEBRUARY, APRIL, AND JULY

Results of a statistical analysis of the first-guess error field for the months of February 1991 and April 1991 are presented in Jacobs and Gaver (1991).

In this section we report results concerning the use of models fit with July data (respectively February or April) to predict February or April (respectively July) mean square first-guess error. These results give an indication of the possibility of using a model fit with one month's data to predict another month's data.

4.1 Observed Wind Covariate Models

In this subsection we report results for normal models (1)–(3) using observed wind components as covariates. There are six analyses; one for the u-wind component (respectively v-wind component) for each pressure level.

Table 7 shows the values of the parameter estimates and estimated standard errors for the February, April, and July data. The minor discrepancies with the values in Jacobs and Gaver (1991) are due to the deletion of the suspect 0 wind values from the data sets in this analysis. Table 8 shows the values of $\tilde{\ell}$ for February data (respectively July data) using parameters fit using February data (respectively July data). Values of $\tilde{\ell}$ are also presented for February (respectively July) data using parameters fit using July (respectively February) data. Once again, larger values of $\tilde{\ell}$ indicate better model fit. The underlined value in each row is the maximum value in that row. The bold italicized value in each row is the maximum value of $\tilde{\ell}$ for the two one-variate models.

The values of $\tilde{\ell}$ for February data (respectively July data) using parameters fit using July data (respectively February data) are maximized by the two-variate model in all but one case; between the two one-variate models $\tilde{\ell}$ is the maximized for the model involving $s(t)$ except in 3 cases.

Comparing the value of $\tilde{\ell}$, $\tilde{\ell}_c$, for the model of constant variance (5) for February (respectively July) data using parameters estimated from February (respectively July) data with that for the prediction value of $\tilde{\ell}$ for the models (2)–(3) for February (respectively July) data using parameters estimated from July (respectively February) data indicate the following. The values of $\tilde{\ell}$ for

TABLE 7. PARAMETER ESTIMATES (STANDARD ERRORS) OBSERVED WIND COVARIATES

Pressure Level	Wind Comp.	Data Set	One-Variate Models $r(t)$ α	β	$s(t)$ α	β	Two-variate Models $\log MSE = \alpha+\beta_1 r(t)+\beta_2 s(t)$ α	β_1	β_2
850	u	July	1.45	0.11	1.46	0.09	1.20	0.08	0.06
			(0.04)	(0.006)	(0.04)	(0.005)	(0.05)	(0.007)	(0.006)
		Apr.	1.86	0.09	1.68	0.009	1.42	0.05	0.07
			(0.04)	(0.005)	(0.04)	(0.004)	(0.05)	(0.005)	(0.005)
		Feb.	2.09	0.05	1.92	0.05	1.74	0.03	0.04
			(0.04)	(0.005)	(0.05)	(0.004)	(0.05)	(0.005)	(0.004)
	v	July	1.55	0.11	1.53	0.09	1.29	0.08	0.06
			(0.04)	(0.006)	(0.04)	(0.006)	(0.05)	(0.007)	(0.006)
		Apr.	1.84	0.09	1.68	0.09	1.43	0.06	0.07
			(0.04)	(0.005)	(0.04)	(0.005)	(0.05)	(0.005)	(0.005)
		Feb.	2.15	0.05	1.71	0.07	1.50	0.03	0.06
			(0.04)	(0.004)	(0.05)	(0.004)	(0.05)	(0.005)	(0.004)
500	u	July	1.40	0.11	1.58	0.05	1.17	0.10	0.03
			(0.04)	(0.006)	(0.04)	(0.003)	(0.05)	(0.006)	(0.004)
		Apr.	2.12	0.06	2.22	0.03	1.81	0.05	0.02
			(0.04)	(0.003)	(0.04)	(0.002)	(0.05)	(0.003)	(0.002)
		Feb.	2.22	0.05	2.40	0.02	2.02	0.05	0.01
			(0.04)	(0.003)	(0.05)	(0.002)	(0.05)	(0.003)	(0.002)
	v	July	1.49	0.10	1.66	0.04	1.27	0.09	0.03
			(0.04)	(0.006)	(0.05)	(0.004)	(0.05)	(0.007)	(0.004)
		Apr.	2.03	0.06	1.99	0.04	1.68	0.05	0.02
			(0.04)	(0.003)	(0.04)	(0.002)	(0.05)	(0.004)	(0.003)
		Feb.	2.28	0.04	2.32	0.02	2.02	0.04	0.01
			(0.04)	(0.003)	(0.05)	(0.002)	(0.05)	(0.003)	(0.002)
250	u	July	2.42	0.06	2.52	0.03	2.13	0.05	0.02
			(0.04)	(0.003)	(0.04)	(0.002)	(0.05)	(0.004)	(0.002)
		Apr.	2.76	0.04	2.67	0.03	2.30	0.04	0.02
			(0.04)	(0.002)	(0.04)	(0.001)	(0.05)	(0.002)	(0.002)
		Feb.	3.02	0.04	2.56	0.03	2.28	0.03	0.03
			(0.04)	(0.002)	(0.04)	(0.001)	(0.06)	(0.002)	(0.001)
	v	July	2.44	0.06	2.43	0.03	2.12	0.05	0.02
			(0.04)	(0.003)	(0.04)	(0.002)	(0.05)	(0.008)	(0.002)
		Apr.	2.75	0.04	2.64	0.03	2.28	0.03	0.02
			(0.04)	(0.002)	(0.04)	(0.001)	(0.05)	(0.002)	(0.002)
		Feb.	3.00	0.003	2.43	0.03	2.23	0.02	0.03
			(0.04)	(0.002)	(0.05)	(0.001)	(0.06)	(0.002)	(0.001)

$$r(t) = [((u(t) - u(t-1))^2 + (v(t) - v(t-1))^2)]^{1/2}$$

$$s(t) = [u(t)^2 + v(t)^2]^{1/2}$$

TABLE 8. NORMAL MODELS
VALUES OF LOG-LIKELIHOOD
OBSERVED WIND COVARIATES
FEBRUARY AND JULY

Pressure Level	Wind Comp.	Data Set	Model	Constant	One-Variate Models $r(t)$	One-Variate Models $s(t)$	Two-variate Models
850	u	July	July	–11269.3	***–10787.3***	–10849.5	–10695.2
		Feb.	Feb.	–13211.6	–13071.2	***–13023.0***	–12964.0
		Feb.	July	–13405.8	–13325.2	***–13138.1***	–13116.5
		July	Feb.	–11417.3	–11017.6	***–10963.0***	–10826.0
	v	July	July	–11373.1	***–10992.0***	–11042.4	–10902.7
		Feb.	Feb.	–13333.7	–13204.4	**–12992.0**	–12957.9
		Feb.	July	–13531.8	–13446.7	**–13018.8**	–13078.6
		July	Feb.	–11523.8	–11200.2	***–11059.6***	–10972.4
500	u	July	July	–12205.2	***–11734.6***	–11953.8	–11670.9
		Feb.	Feb.	–16273.0	***–15924.6***	–16151.7	–15892.9
		Feb.	July	–17497.3	***–16399.1***	–16512.5	–16216.4
		July	Feb.	–12913.9	***–12174.5***	–12419.9	–12014.1
	v	July	July	–12221.4	***–11905.2***	–12056.9	–11855.2
		Feb.	Feb.	–15966.1	***–15750.5***	–15859.9	–15707.1
		Feb.	July	–16900.4	–16168.5	***–16066.0***	–15997.7
		July	Feb.	–12790.9	***–12281.1***	–12361.7	–12103.1
250	u	July	July	–15876.6	***–15440.6***	–15592.7	–15335.2
		Feb.	Feb.	–18771.3	–17773.0	***–17619.9***	–17413.1
		Feb.	July	–20206.9	–18045.1	***–17657.2***	–17530.2
		July	Feb.	–16742.6	–15713.3	***–15609.6***	–15386.4
	v	July	July	–15792.6	***–15389.4***	–15481.5	–15273.3
		Feb.	Feb.	–18095.0	–17366.6	***–17227.4***	–17062.1
		Feb.	July	–18953.0	–17603.0	***–17227.8***	–17186.8
		July	Feb.	–16366.7	–15608.7	***–15481.6***	–15323.9

models (2) and (3) fit with data from the other month are larger in the majority of the cases than the corresponding values of $\tilde{\ell}_c$ fit with the data of the same month. This suggests that models (2) and (3) fit using data from the other month have some predictive value over a model of constant variance fit using the data that is to be modeled.

Table 9 shows values of $\tilde{\ell}$ for April data (respectively July data) using parameters fit using April data (respectively July data). Values of $\tilde{\ell}$ are also presented for April data (respectively July data) using parameters fit using July data (respectively February data). The underlined value in each row is the maximum value in that row. The bold italicized value in each row is the maximum value of $\tilde{\ell}$ for the two one-variate models.

The values of $\tilde{\ell}$ for April data (respectively July data) using parameters fit using July data (respectively April data) are maximized by the two-variate model in all cases; between the two one-variate models $\tilde{\ell}$ is maximized in all but five cases for the model involving $r(t)$.

Comparing the value of $\tilde{\ell}$, $\tilde{\ell}_c$ for the model of constant variance (5) for April (respectively July) data using parameters estimated from April (respectively July) data with that for the prediction value of $\tilde{\ell}$ for the models (2)-(3) for April (respectively July) data using parameters estimated from July (respectively April) data indicate the following. The values of $\tilde{\ell}$ for models (2) and (3) fit with data from the other month are larger in the majority of the cases than the corresponding values of $\tilde{\ell}_c$ fit with the data of the same month. This suggests that models (2) and (3) fit using data from the other month have some predictive value over a model of constant variance fit using the data that is to be modeled.

TABLE 9. OBSERVED WIND COVARIATES VALUE OF LOG-LIKELIHOOD APRIL AND JULY

Pressure Level	Wind Comp.	Data Set	Model	Constant	One-Variate Models $r(t)$	One-Variate Models $s(t)$	Two-variate Models
850	u	July	July	–11269.3	***–10787.3***	–10849.5	–10695.2
		April	April	–13814.3	–13460.9	***–13313.6***	–13205.7
		April	July	–14081.0	–13592.6	***–13369.6***	–13265.6
		July	April	–11460.5	–10911.4	***–10901.0***	–10743.5
	v	July	July	–11373.1	***–10992.0***	–11042.4	–10902.7
		April	April	–13837.2	–13421.2	***–13389.5***	–13229.7
		April	July	–14067.2	–13490.0	***–13423.9***	–13251.4
		July	April	–11540.6	***–11058.4***	–11073.6	–10920.7
500	u	July	July	–12205.2	***–11734.6***	–11953.8	–11670.9
		April	April	–16262.1	***–15875.3***	–16055.1	–15775.5
		April	July	–17101.2	***–16259.4***	–16391.3	–16020.9
		July	April	–12714.1	***–12074.3***	–12272.1	–11893.0
	v	July	July	–12221.4	***–11905.2***	–12056.9	–11855.2
		April	April	–16476.6	***–15698.2***	–15843.3	–15584.2
		April	July	–17472.3	***–15913.3***	–16008.0	–15703.2
		July	April	–12807.2	***–12095.5***	–12198.1	–11946.8
250	u	July	July	–15876.6	***–15440.6***	–15592.7	–15335.2
		April	April	–20104.9	***–17863.0***	–18119.6	–17705.0
		April	July	–21601.8	***–17954.3***	–18144.5	–17750.3
		July	April	–16723.4	***–15514.5***	–15619.0	–15357.4
	v	July	July	–15792.6	***–15389.4***	–15481.5	–15273.3
		April	April	–18674.8	***–17610.7***	–17853.7	–17473.4
		April	July	–19096.9	***–17691.2***	–17884.1	–17525.5
		July	April	–16089.6	***–15448.2***	–15507.4	–15296.4

Table 10 shows the fraction of increase in $\tilde{\ell}$ of using a model with parameters estimated using another month to predict variance in the current month compared to using the best constant variance model fit with the current month. The results suggest that the models for other months do have some predictive ability. Models fit using April data appear to have more predictive ability for July than those fit using February data. The predictive ability appears greater at the 250 mb level.

4.2 First-guess Wind Covariate Models

In this section we report results for normal models (1)–(3) using first-guess wind components as covariates.

Table 11 shows the values of the parameter estimates and standard errors for February data, April data and July data. The minor discrepancies with values reported in Jacobs and Gaver (1991) are due to the deletion of suspicious 0 wind values from the data sets. Table 12 shows the values of $\tilde{\ell}$ for February data (respectively April data) using parameters estimated from February data (respectively July data). Values of $\tilde{\ell}$ are also presented for February data (respectively July data) using parameters estimated from July data (respectively February data). The underlined value in each row is the maximum value in that row. The bold italicized value in each row is the maximum value of $\tilde{\ell}$ for the two one-variate models.

The values of $\tilde{\ell}$ for the observed wind covariates are larger than those for the first-guess wind covariates in all cases. This suggests that the observed wind covariates provide better models of the data both in terms of goodness-of-fit and prediction.

TABLE 10. FRACTION OF INCREASE IN LOG-LIKELIHOOD

$$(\tilde{\ell}-\tilde{\ell}_c)/|\tilde{\ell}_c|$$

Pressure Level	Wind Comp.	Data Set	Model	One-variate Models $r(t)$	$s(t)$	Two-variate Models
850	u	July	Feb.	0.02	0.03	0.04
		July	Apr.	0.03	0.03	0.05
		Feb.	July	*	0.006	0.007
		Apr.	July	0.02	0.03	0.04
	v	July	Feb.	0.02	0.03	0.04
		July	Apr.	0.03	0.03	0.04
		Feb.	July	*	0.006	0.007
		Apr.	July	0.03	0.03	0.04
500	u	July	Feb.	0.003	*	0.02
		July	Apr.	0.01	*	0.03
		Feb.	July	*	*	*
		Apr.	July	0.00	*	0.01
	v	July	Feb.	*	*	0.01
		July	Apr.	0.01	0.002	0.02
		Feb.	July	*	*	*
		Apr.	July	0.03	0.03	0.05
250	u	July	Feb.	0.01	0.02	0.03
		July	Apr.	0.02	0.02	0.03
		Feb.	July	0.04	0.06	0.07
		Apr.	July	0.11	0.10	0.12
	v	July	Feb.	0.01	0.02	0.03
		July	Apr.	0.02	0.02	0.03
		Feb.	July	0.03	0.05	0.05
		Apr.	July	0.05	0.04	0.06

*: $\tilde{\ell}_c$ (data described by model with constant variance estimated using same data)

$> \tilde{\ell}$ (data described by model fit using other month)

TABLE 11. FIRST GUESS WIND COVARIATES PARAMETER ESTIMATES (STANDARD ERRORS)

			One-variate Models				Two-variate Models		
Pressure Level	Wind Comp.	Data Set	$r(t)$ α	$r(t)$ β	$s(t)$ α	$s(t)$ β	$\log \text{MSE} = \alpha+\beta_1 r(t)+\beta_2 s(t)$ α	β_1	β_2
850	u	July	2.00	0.05	2.07	0.02	1.96	0.045	0.009
			(0.04)	(0.009)	(0.04)	(0.006)	(0.05)	(0.009)	(0.007)
		Apr	2.37	0.03	2.12	0.05	2.11	0.003	0.045
			(0.04)	(0.006)	(0.04)	(0.004)	(0.05)	(0.007)	(0.004)
		Feb	2.47	0.01	2.25	0.03	2.27	–0.01	–0.03
			(0.04)	(0.005)	(0.04)	(0.004)	(0.05)	(0.005)	(0.004)
850	v	July	1.98	0.06	1.94	0.04	1.86	0.04	0.03
			(0.04)	(0.008)	(0.04)	(0.006)	(0.05)	(0.009)	(0.007)
		Apr	2.46	0.02	2.21	0.04	2.22	–0.002	0.04
			(0.04)	(0.006)	(0.04)	(0.004)	(0.05)	(0.007)	(0.004)
		Feb	2.45	0.01	2.35	0.02	2.34	0.003	0.02
			(0.04)	(0.005)	(0.04)	(0.003)	(0.04)	(0.005)	(0.004)
500	u	July	1.93	0.05	1.82	0.03	1.72	0.03	0.03
			(0.04)	(0.007)	(0.04)	(0.003)	(0.05)	(0.008)	(0.004)
		Apr	2.51	0.03	2.25	0.03	2.14	0.02	0.03
			(0.04)	(0.005)	(0.04)	(0.002)	(0.05)	(0.005)	(0.002)
		Feb	2.61	0.03	2.54	0.02	2.38	0.02	0.01
			(0.04)	(0.004)	(0.05)	(0.002)	(0.05)	(0.004)	(0.002)
500	v	July	1.97	0.04	1.92	0.02	1.83	0.03	0.02
			(0.04)	(0.007)	(0.04)	(0.004)	(0.05)	(0.007)	(0.004)
		Apr	2.33	0.06	1.96	0.05	1.76	0.03	0.04
			(0.04)	(0.005)	(0.04)	(0.002)	(0.05)	(0.005)	(0.002)
		Feb	2.71	0.01	2.47	0.02	2.44	0.004	0.01
			(0.04)	(0.004)	(0.05)	(0.002)	(0.05)	(0.004)	(0.002)
250	u	July	2.90	0.03	2.79	0.02	2.70	0.02	0.02
			(0.04)	(0.004)	(0.04)	(0.002)	(0.05)	(0.005)	(0.002)
		Apr	4.01	–0.01	3.48	0.01	3.63	–0.02	0.02
			(0.04)	(0.004)	(0.05)	(0.002)	(0.06)	(0.004)	(0.002)
		Feb	3.67	0.02	2.94	0.03	2.75	0.02	0.03
			(0.04)	(0.003)	(0.05)	(0.002)	(0.06)	(0.03)	(0.001)
250	v	July	2.79	0.04	2.71	0.02	2.55	0.03	0.02
			(0.04)	(0.004)	(0.04)	(0.002)	(0.05)	(0.004)	(0.002)
		Apr	3.27	0.03	2.80	0.03	2.68	0.01	0.02
			(0.04)	(0.003)	(0.05)	(0.002)	(0.05)	(0.003)	(0.002)
		Feb	3.48	0.02	3.08	0.02	2.84	0.02	0.02
			(0.04)	(0.003)	(0.05)	(0.002)	(0.06)	(0.003)	(0.001)

TABLE 12. FIRST-GUESS WIND COVARIATES
VALUE OF LOG-LIKELIHOOD
JULY AND FEBRUARY

Pressure Level	Wind Comp.	Data Set	Model	Constant	One-variate Models $r_f(t)$	One-variate Models $s_f(t)$	Two-variate Models
850	u	July	July	–11879.1	***–11844.0***	–11867.3	–11842.0
		Feb.	Feb.	–14370.0	–14367.8	***–14316.7***	–14315.3
		Feb.	July	–14597.0	–14553.6	***–14429.2***	–14492.6
		July	Feb.	–12046.1	–12022.4	***–11942.7***	–11954.1
850	v	July	July	–11925.2	***–11873.3***	–11875.4	–11853.2
		Feb.	Feb.	–14399.2	–14392.7	***–14373.5***	–14373.1
		Feb.	July	–14618.3	–14563.6	***–14487.1***	–14504.7
		July	Feb.	–12087.0	–12042.5	***–11997.3***	–11991.0
500	u	July	July	–13308.2	–13255.8	***–13223.8***	–13203.8
		Feb.	Feb.	–17944.6	***–17889.7***	–17890.6	–17853.6
		Feb.	July	–19419.3	–18592.2	***–18411.8***	–18170.3
		July	Feb.	–14143.8	–13857.6	***–13788.7***	–13633.3
500	v	July	July	–13314.7	***–13274.6***	–13274.8	–13253.6
		Feb.	Feb.	–17592.4	–17587.0	***–17541.6***	–17540.5
		Feb.	July	–18727.5	–18262.2	***–17994.9***	–17890.6
		July	Feb.	–13992.0	–13909.7	***–13684.2***	–13659.9
250	u	July	July	–17182.5	–17117.0	***–17080.6***	–17059.0
		Feb.	Feb.	–20872.6	–20836.3	***–20538.2***	–20505.4
		Feb.	July	–22345.1	–21676.2	***–21016.2***	–20887.3
		July	Feb.	–18057.7	–17829.7	***–17266.9***	–17178.3
250	v	July	July	–17005.0	–16900.6	***–16889.1***	–16833.3
		Feb.	Feb.	–20075.0	–20031.7	***–19925.2***	–19876.6
		Feb.	July	–20975.0	–20485.3	***–20131.7***	–19988.6
		July	Feb.	–17593.9	–17376.1	***–17090.4***	–16944.4

The values of $\tilde{\ell}$ for February data (respectively July data) using parameters fit using July data (respectively February data) are maximized most of the time by the two-variate model.

A comparison of the value of $\tilde{\ell}$, $\tilde{\ell}_c$, for the constant variance model of February (respectively July) data fit using the same month February (respectively July) data and the prediction values of $\tilde{\ell}$ for models (1)–(3) of February (respectively July) data using parameters estimated from the other month of July (respectively February) indicate the following. A majority of the time $\tilde{\ell}_c$ is larger than the corresponding values of $\tilde{\ell}$ for models (1)–(3) fit with the other month's data. This suggests that the first-guess covariate models fit using the other month's data may not describe the data as well as a constant variance model fit using the data being modeled. This may be an indication that models fit using first-guess February wind (respectively July wind) data are not good predictors of July (respectively February) wind component error.

Table 13 presents values of $\tilde{\ell}$ similar to those of Table 12 except that they are for the months of April and July. Comparison of the values of $\tilde{\ell}_c$ for data of one month fit with a constant variance model using the same data and the corresponding value of $\tilde{\ell}$ for the data using models with parameters estimated using the other month suggests that models using first-guess covariates do not have much predictive ability across these months. Table 14 presents the fraction of increase $(\tilde{\ell}-\tilde{\ell}_c)/|\tilde{\ell}_c|$ for the models with first guess covariates. Once again the results suggest that models using first guess wind components do not have much predictive ability across months.

TABLE 13. VALUE OF LOG-LIKELIHOOD FIRST-GUESS WIND COVARIATES APRIL AND JULY

Pressure Level	Wind Comp.	Data Set	Model	Constant	One-variate Models $r_f(t)$	One-variate Models $s_f(t)$	Two-variate Models
850	u	July	July	–11879.1	***–11844.0***	–11867.3	–11842.0
		Apr.	Apr.	–14757.3	–14736.8	***–14626.1***	–14625.9
		Apr.	July	–14999.4	–14901.0	***–14818.1***	–14834.6
		July	Apr.	–12052.2	–11988.1	***–11950.1***	–11945.1
850	v	July	July	–11925.2	***–11873.3***	–11875.4	–11853.2
		Apr.	Apr.	–14949.1	–14937.9	***–14848.1***	–14848.0
		Apr.	July	–15247.5	–15169.4	***–15002.4***	–15022.3
		July	Apr.	–12133.8	–12077.1	***–11992.8***	–11996.6
500	u	July	July	–13308.2	–13255.8	***–13223.8***	–13203.8
		Apr.	Apr.	–17905.8	–17860.4	***–17761.5***	–17742.2
		Apr.	July	–18865.3	–18381.2	***–18190.3***	–18031.5
		July	Apr.	–13883.0	–13686.2	***–13530.8***	–13499.0
500	v	July	July	–13314.7	***–13274.6***	–13274.8	–13253.6
		Apr.	Apr.	–18112.7	–17948.5	***–17703.9***	–17645.9
		Apr.	July	–19233.9	–18557.9	***–18350.3***	–18120.7
		July	Apr.	–13967.8	–13597.8	***–13465.5***	–13371.9
250	u	July	July	–17182.5	-17117.0	***–17080.6***	–17059.0
		Apr.	Apr.	–22104.4	–22091.9	***–22033.1***	–22001.7
		Apr.	July	–23605.6	–23431.0	***–22954.8***	–22967.8
		July	Apr.	–18030.1	–18149.2	***–17715.7***	–17847.7
250	v	July	July	–17005.0	–16900.6	***–16889.1***	–16833.3
		Apr.	Apr.	–20637.7	–20576.6	***–20355.8***	–20336.7
		Apr.	July	–21139.5	–20837.5	***–20503.3***	–20453.8
		July	Apr.	–17346.8	–17149.6	***–16965.5***	–16905.3

TABLE 14. FRACTION OF INCREASE IN LOG-LIKELIHOOD

$(\tilde{\ell}-\tilde{\ell}_c)/|\tilde{\ell}_c|$

FIRST-GUESS WIND COVARIATES

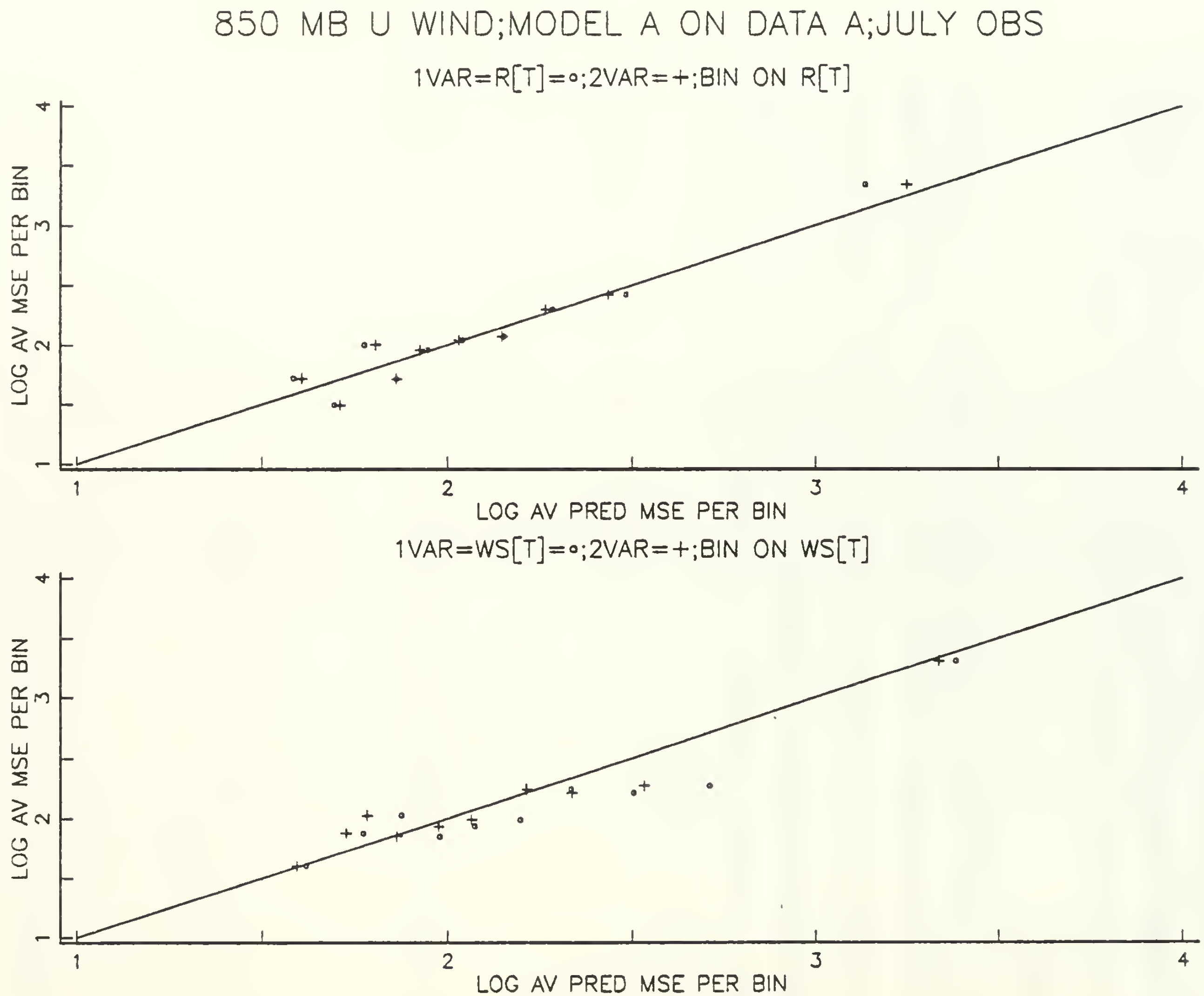
Pressure Level	Wind Comp.	Data Set	Model	One-variate Models $r_f(t)$	$s_f(t)$	Two-variate Models
850	u	July	Feb.	*	*	*
		July	Apr.	*	*	*
		Feb.	July	*	*	*
		Apr.	July	*	*	*
850	v	July	Feb.	*	*	*
		July	Apr.	*	*	*
		Feb.	July	*	*	*
		Apr.	July	*	*	*
500	u	July	Feb.	*	*	*
		July	Apr.	*	*	*
		Feb.	July	*	*	*
		Apr.	July	*	*	*
500	v	July	Feb.	*	*	*
		July	Apr.	*	*	*
		Feb.	July	*	*	*
		Apr.	July	*	*	*
250	u	July	Feb.	*	*	0.00
		July	Apr.	*	*	*
		Feb.	July	*	*	*
		Apr.	July	*	*	*
250	v	July	Feb.	*	*	0.004
		July	Apr.	*	0.002	0.006
		Feb.	July	*	*	0.004
		Apr.	July	*	0.007	0.009

*: $\tilde{\ell}_c$ (data described by model of constant variance fit using same data) $> \tilde{\ell}$ (data described by model fit using the other month)

4.3 Conclusions

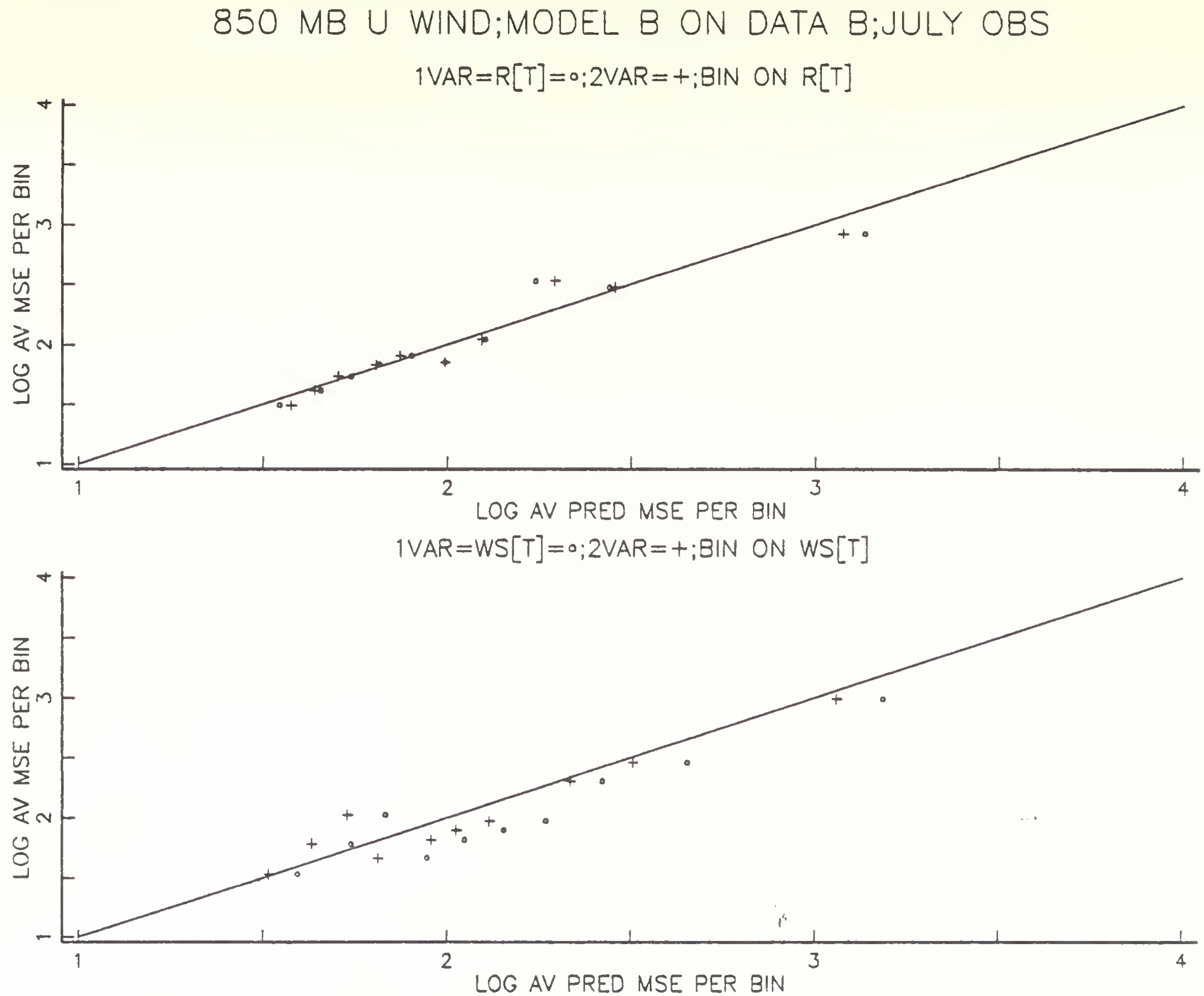
Models (2) and (3) using observed wind components as covariates and fit using February or April (respectively July) data appear to have some predictive value for July (respectively February or April) data. Their predictive ability appears to be better for lower pressure levels. Models fit using April data appear to have more predictive ability than those fit using February data.

Models using first-guess wind covariates do not appear to have predictive ability across these months. It might be that models (1)–(3) fit with first-guess data from other Julys are better predictors of July wind component error. Alternatively, if first-guess winds are to be used as predictors, it might be worthwhile to develop a procedure to update the fitted model parameters using new data as it comes in.



850 MB U WIND;MODEL A ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 1



850 MB U WIND;MODEL B ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 2

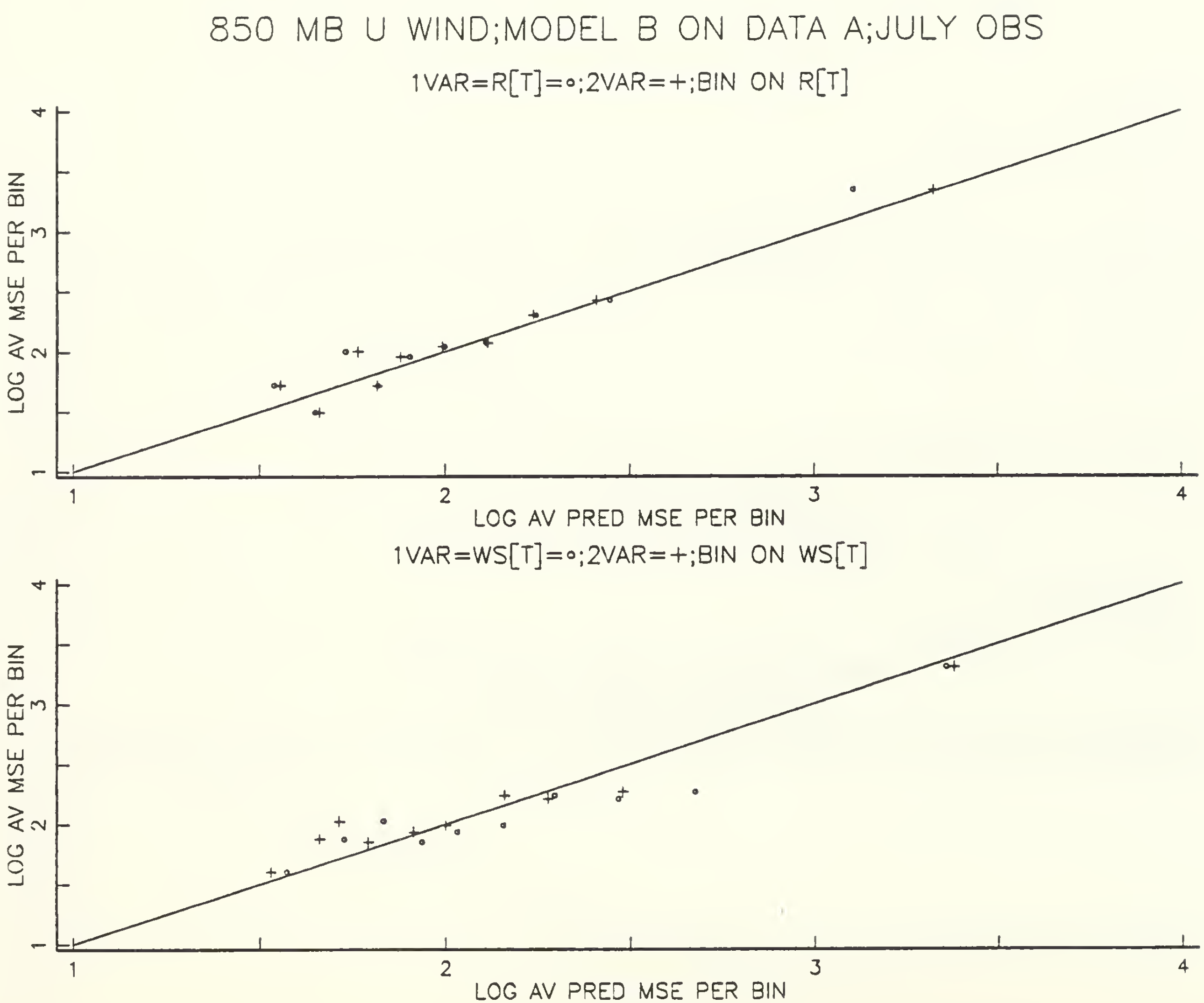

850 MB U WIND;MODEL B ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 3

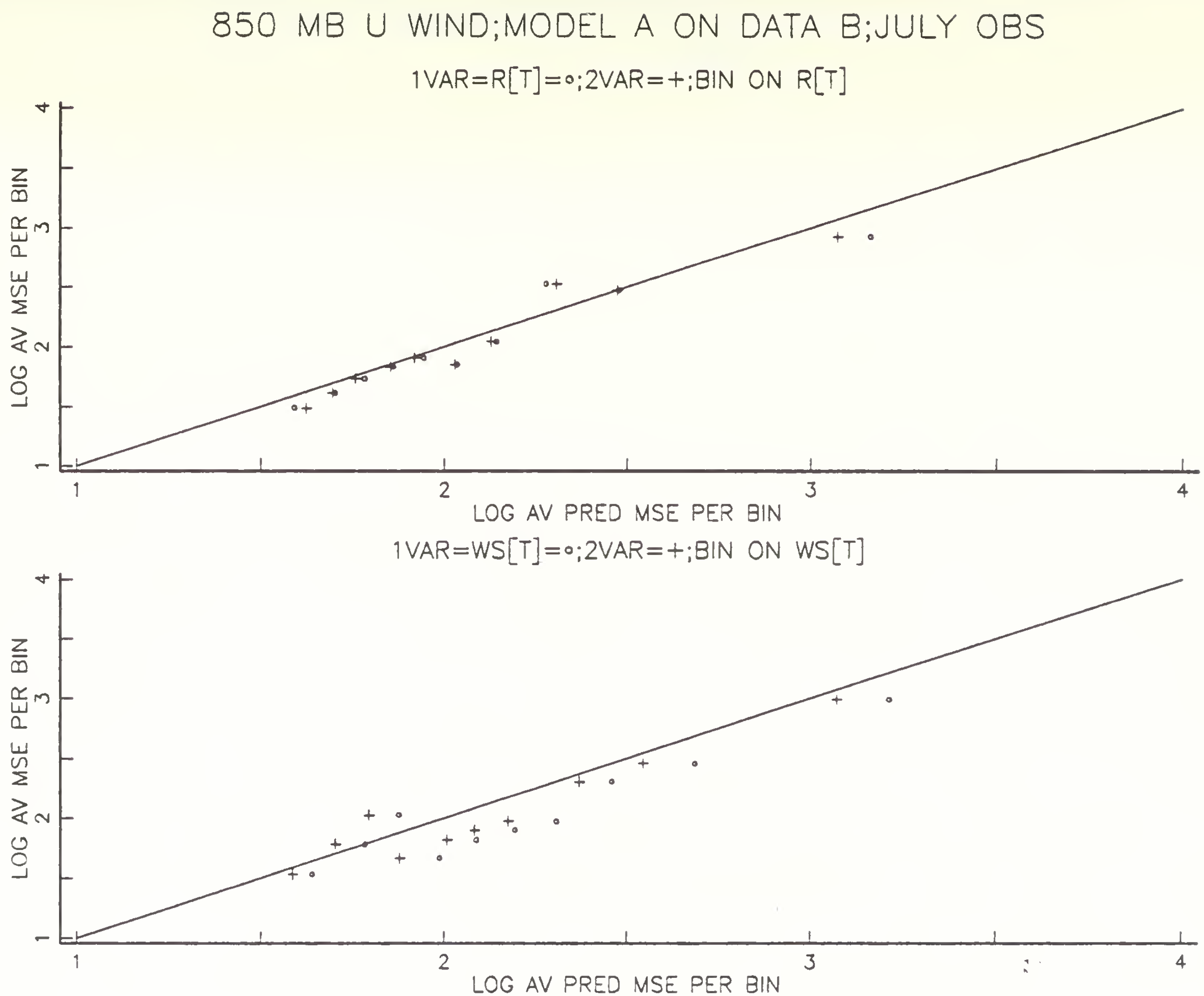

850 MB U WIND;MODEL A ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1
2
3
4

Figure 4

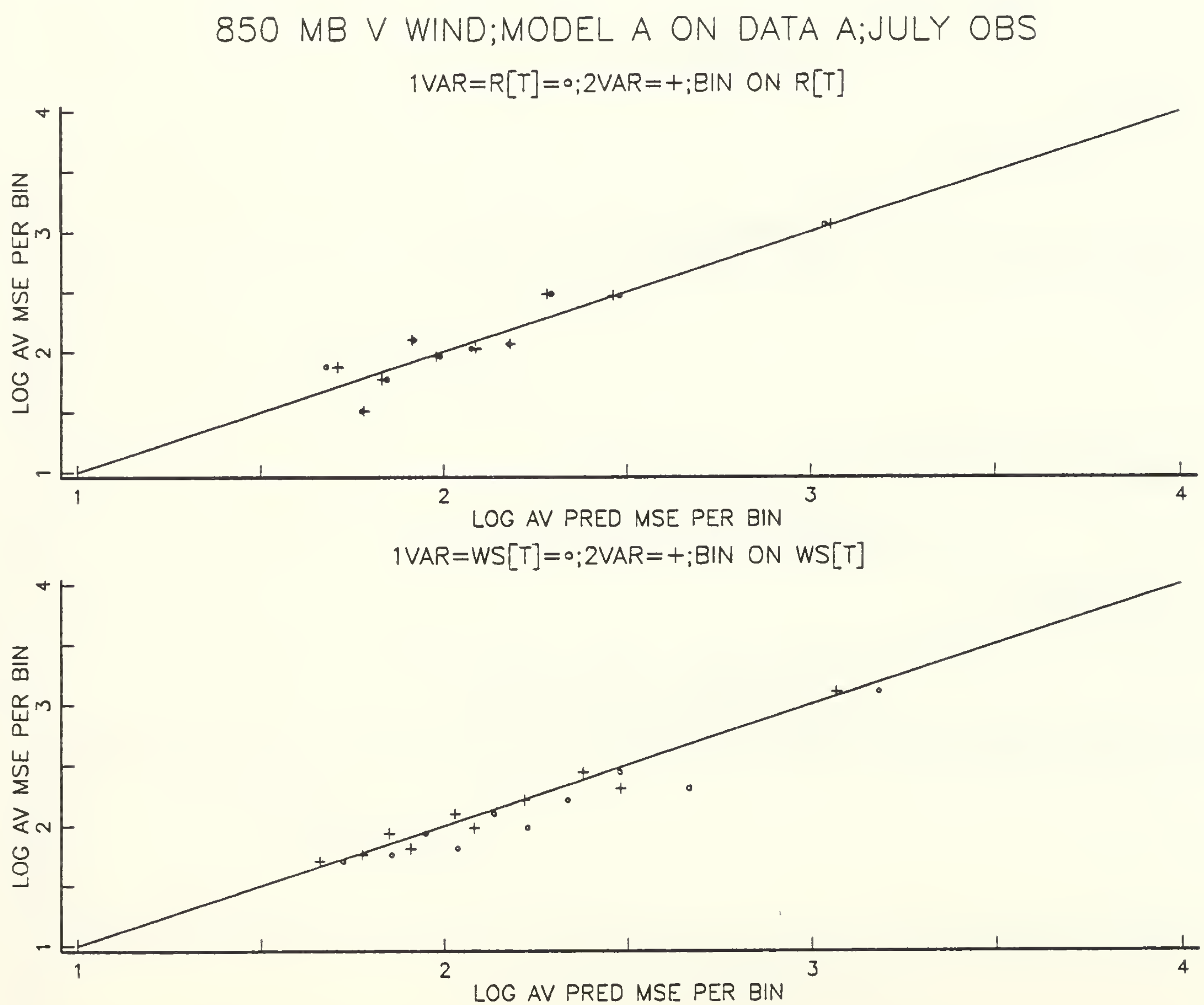

850 MB V WIND;MODEL A ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 5

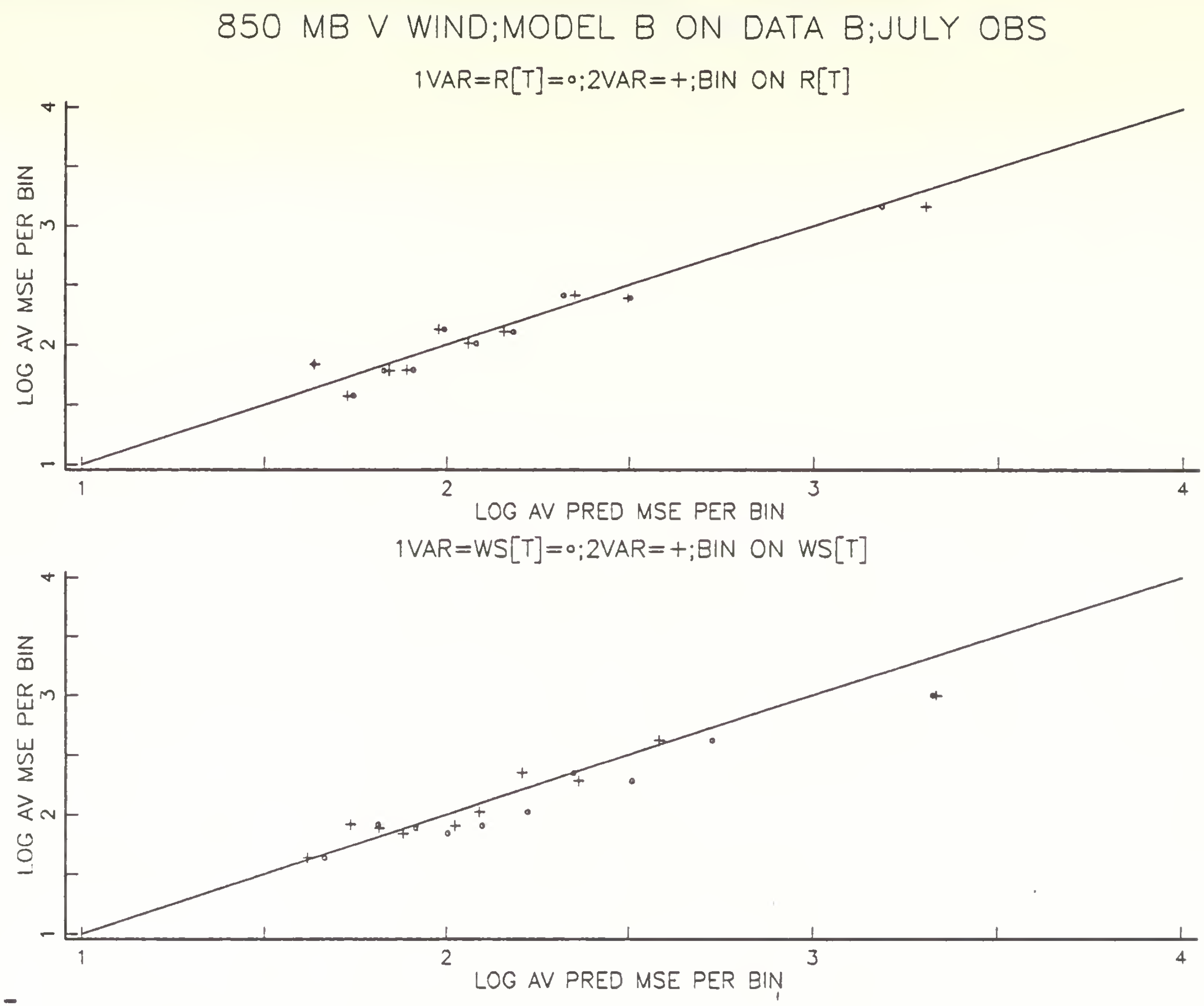
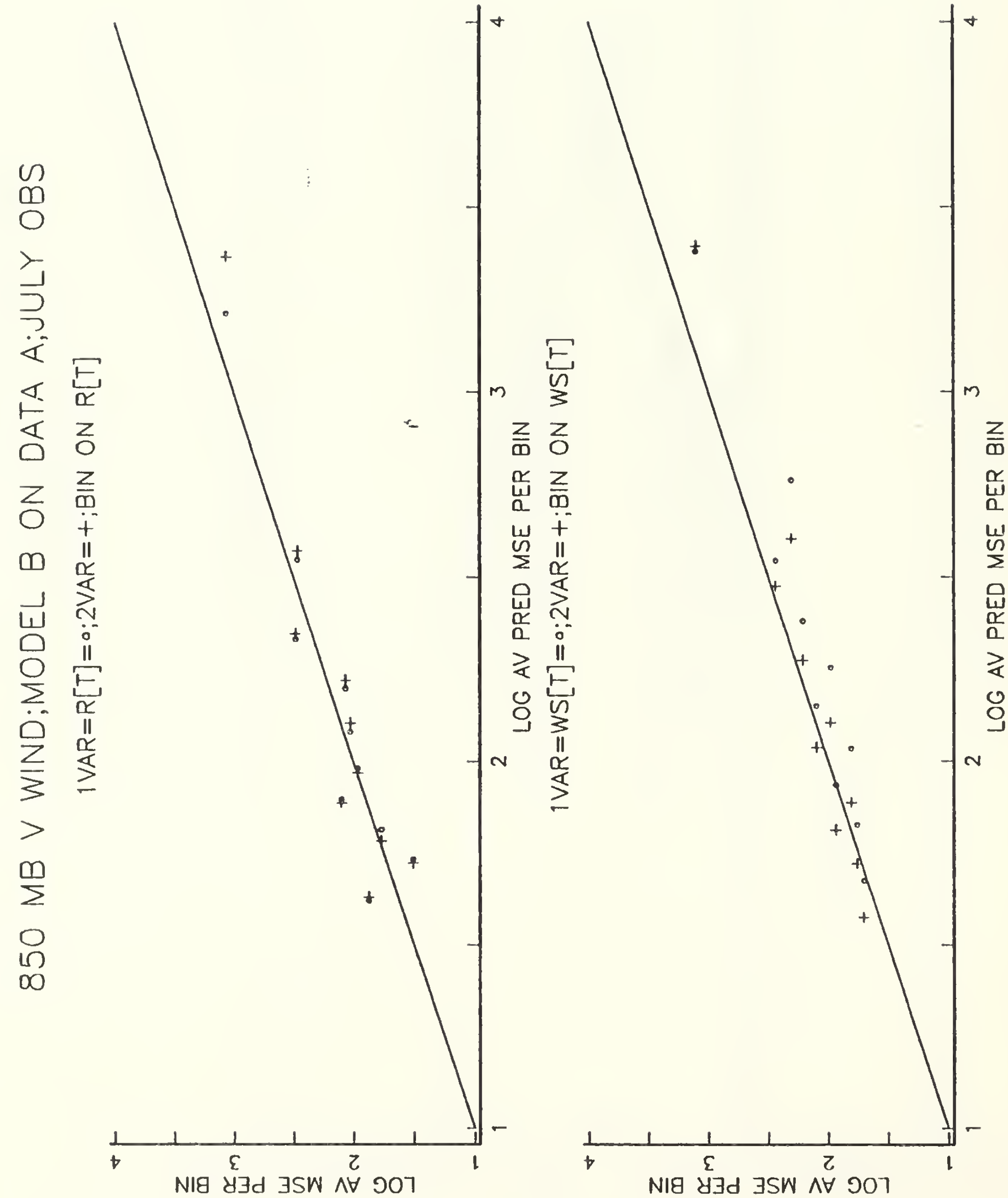
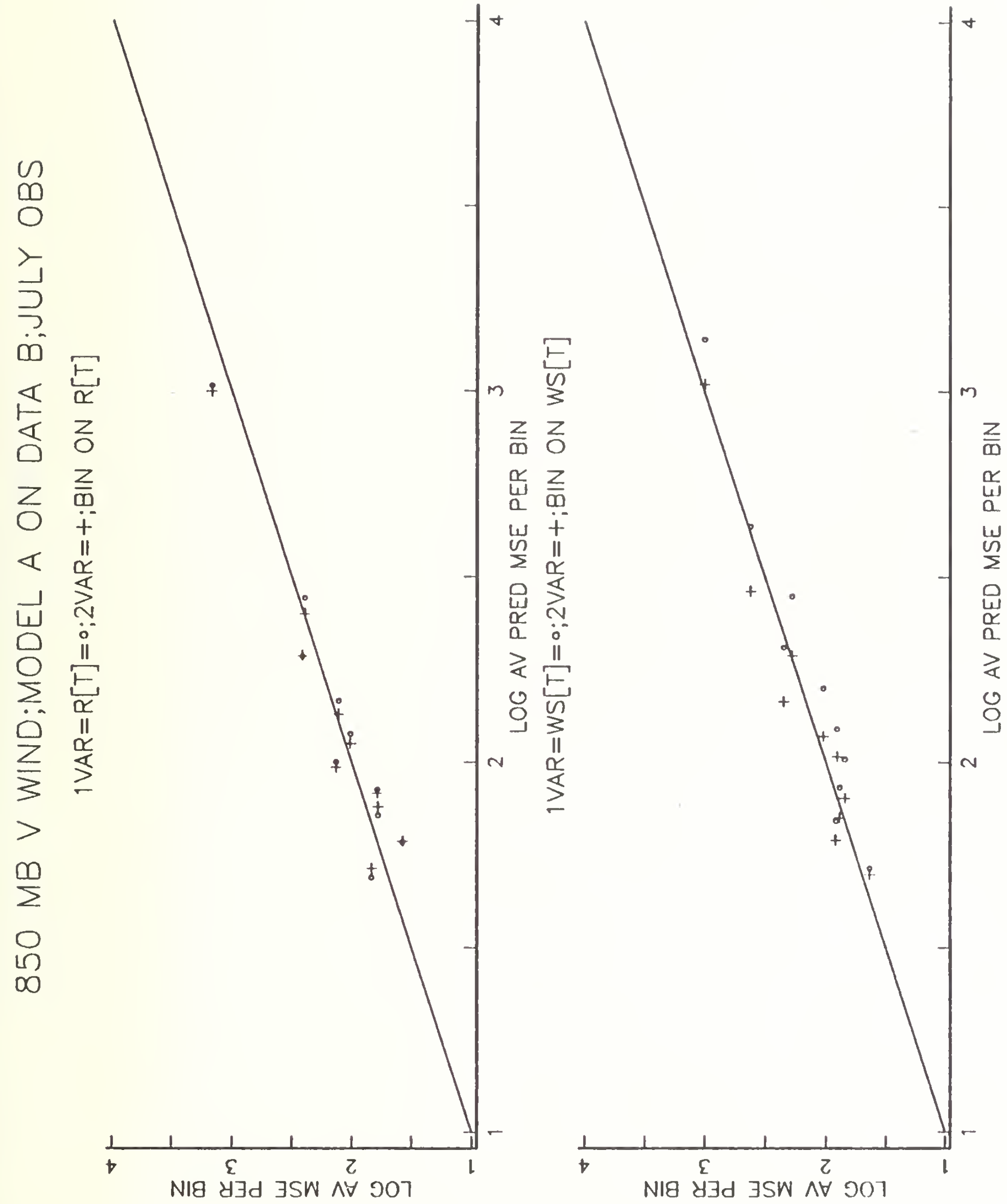

850 MB V WIND;MODEL B ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 6



850 MB V WIND;MODEL B ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 7



850 MB V WIND;MODEL A ON DATA B;JULY OBS
1VAR=R[T]=o;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=o;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1
2
3
4

Figure 8

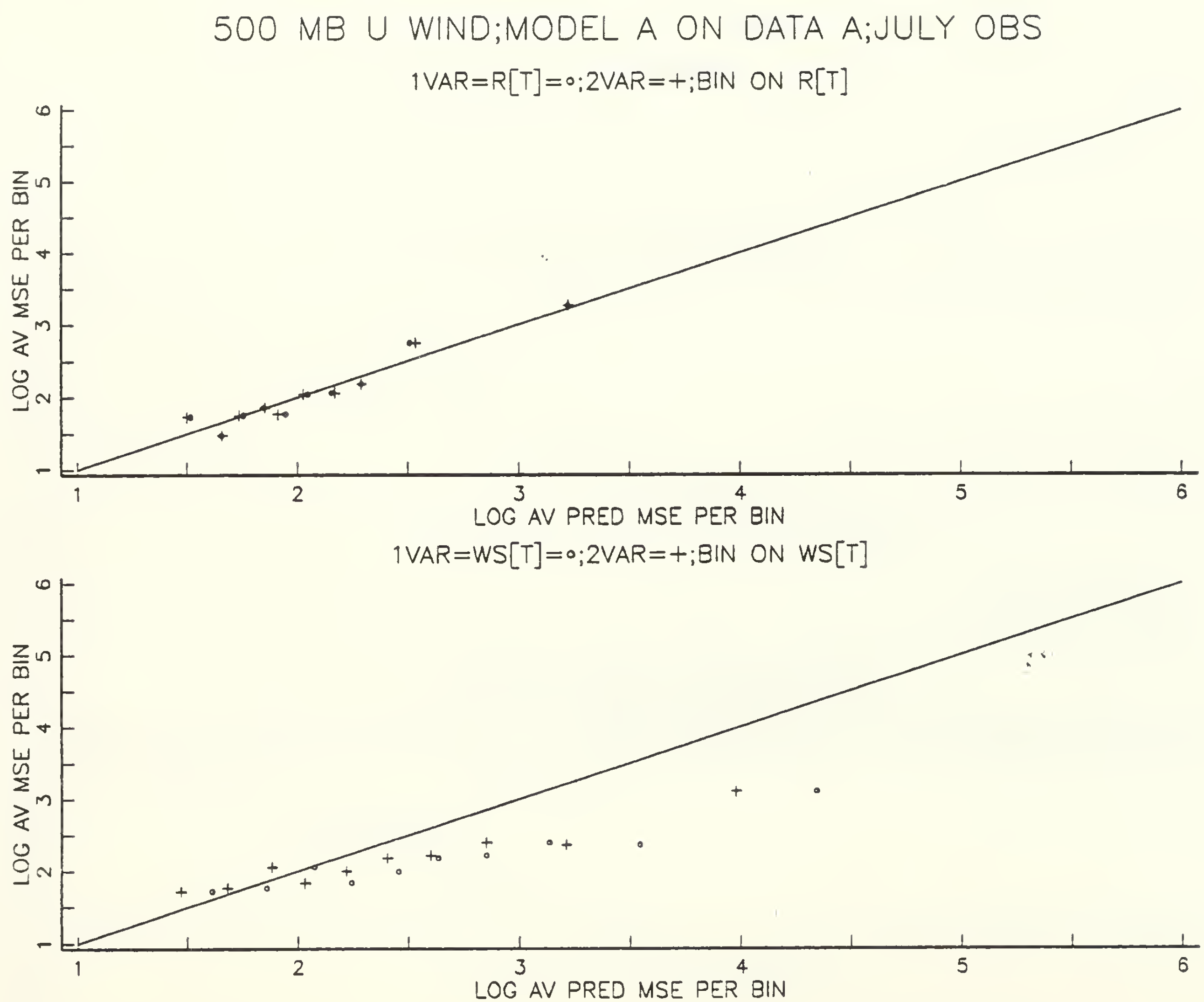

500 MB U WIND;MODEL A ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 9

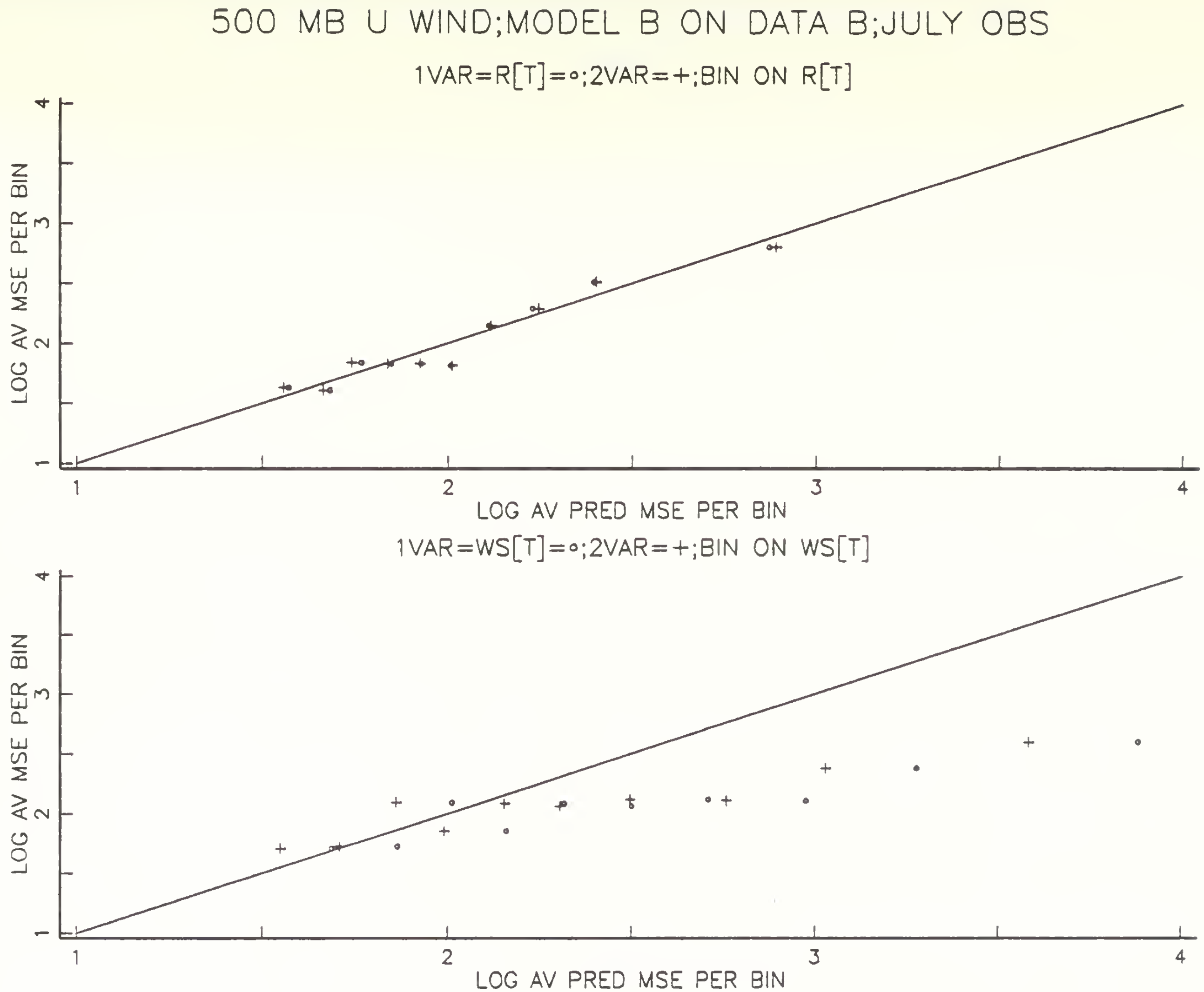

500 MB U WIND;MODEL B ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 10

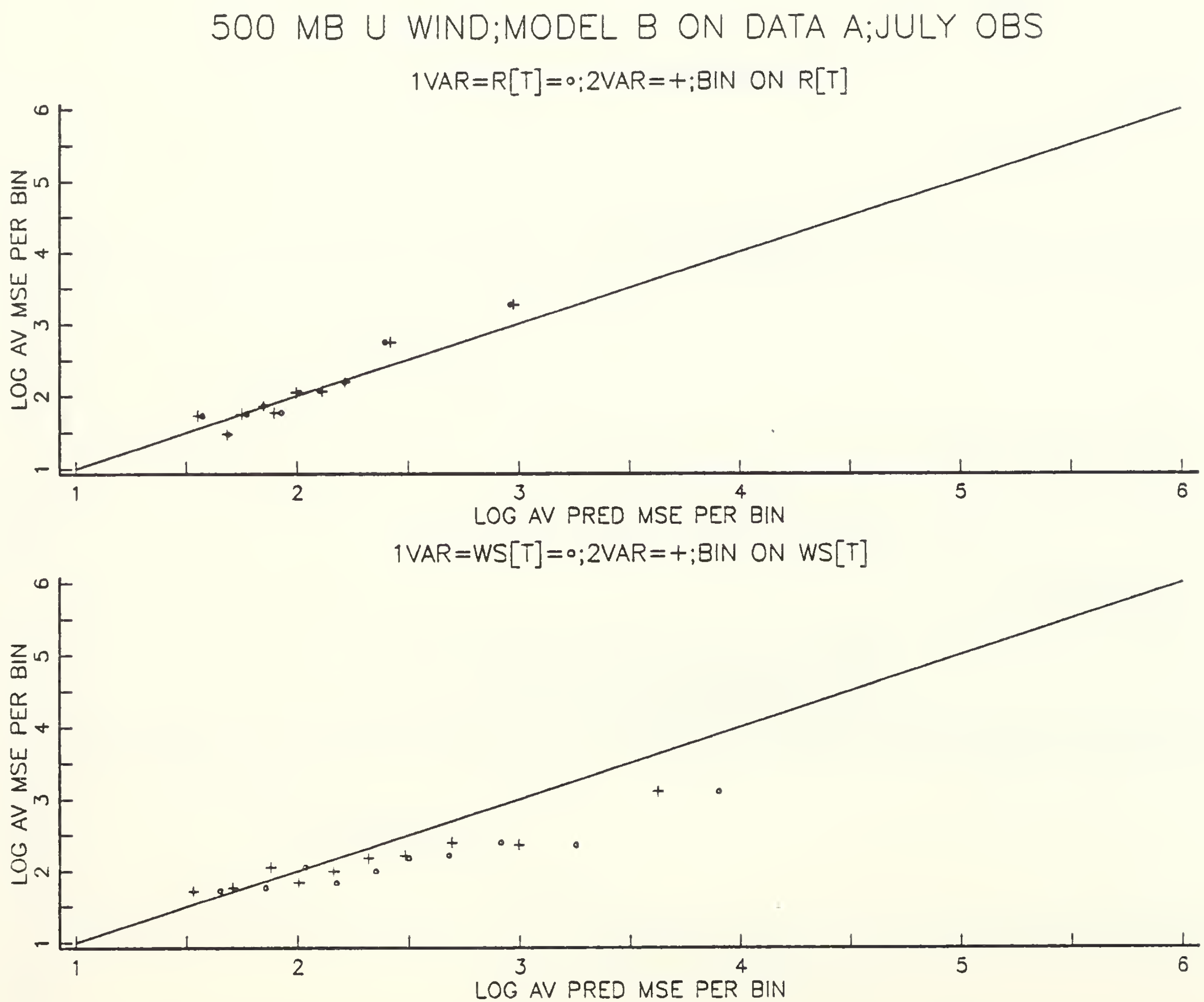
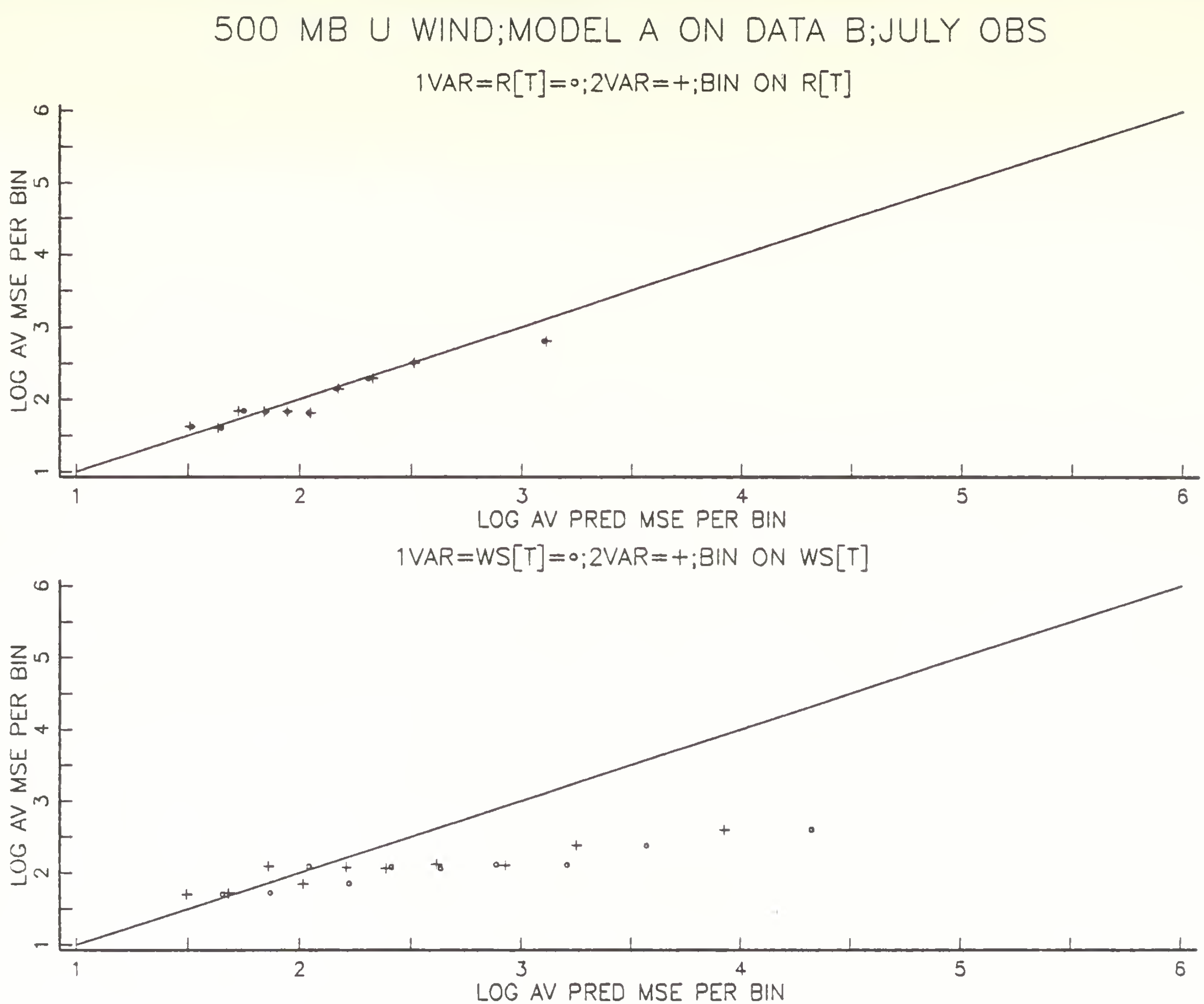

500 MB U WIND;MODEL B ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 11



500 MB U WIND;MODEL A ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 12

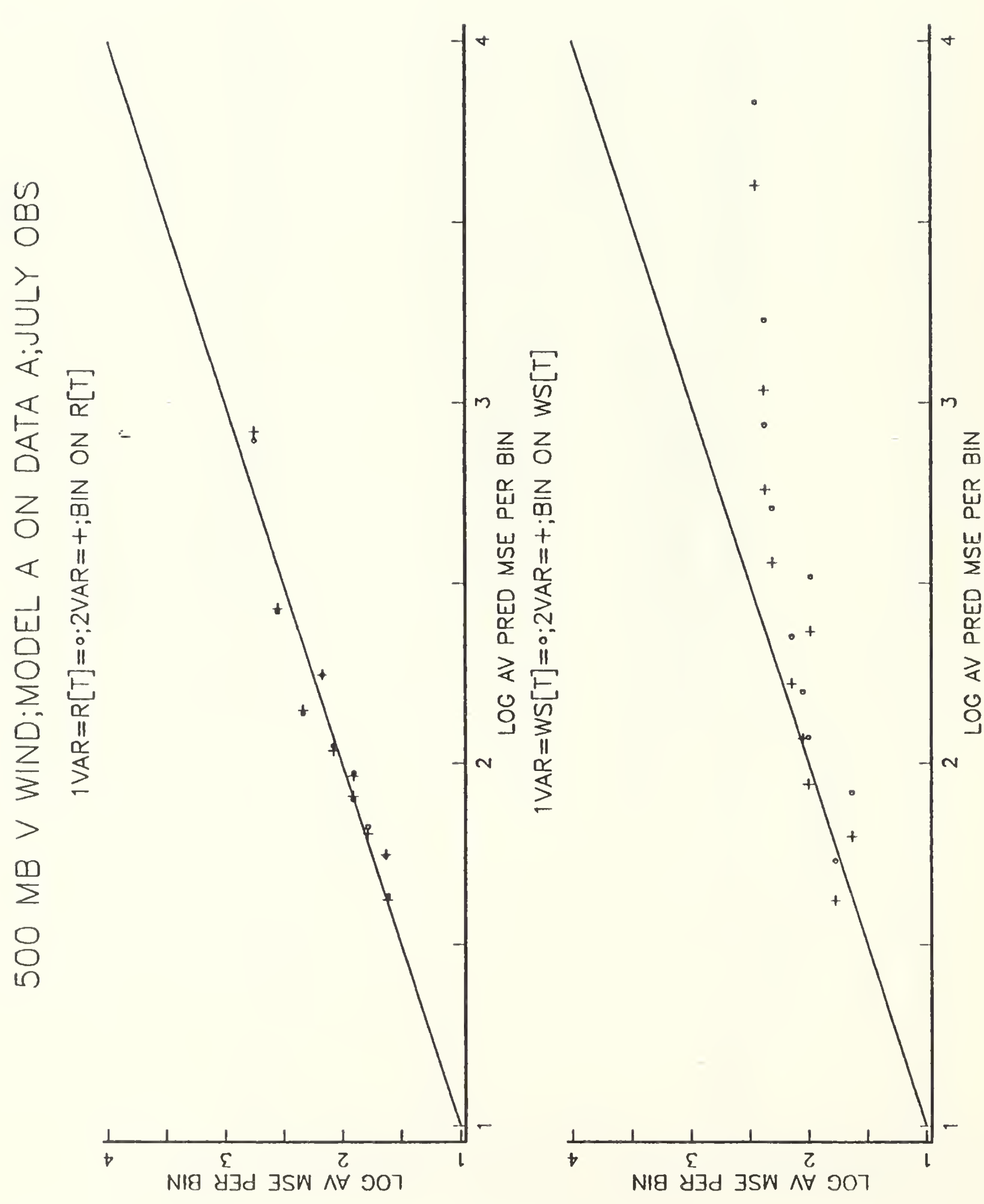
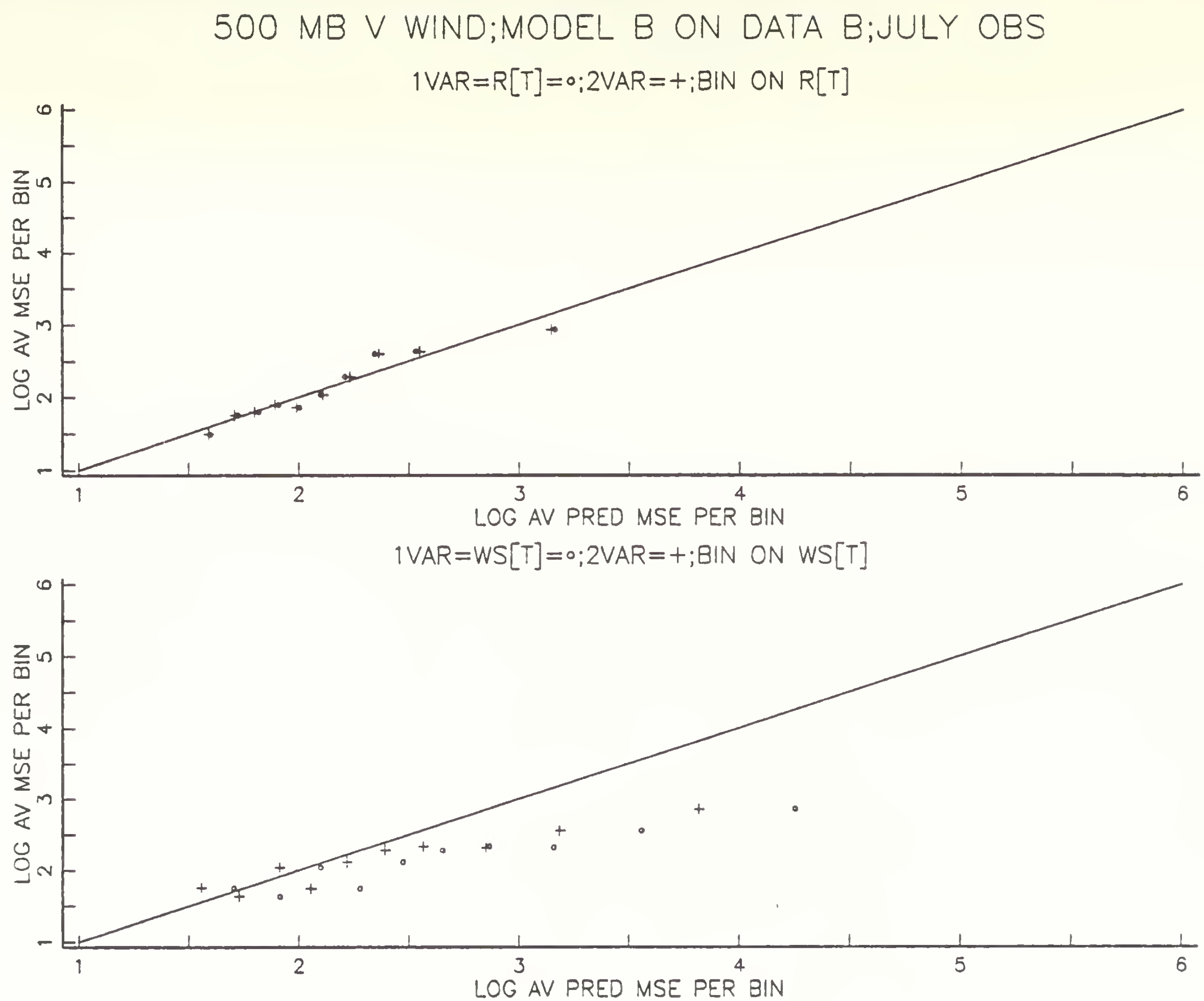

500 MB V WIND;MODEL A ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1
2
3
4

Figure 13



500 MB V WIND;MODEL B ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 14

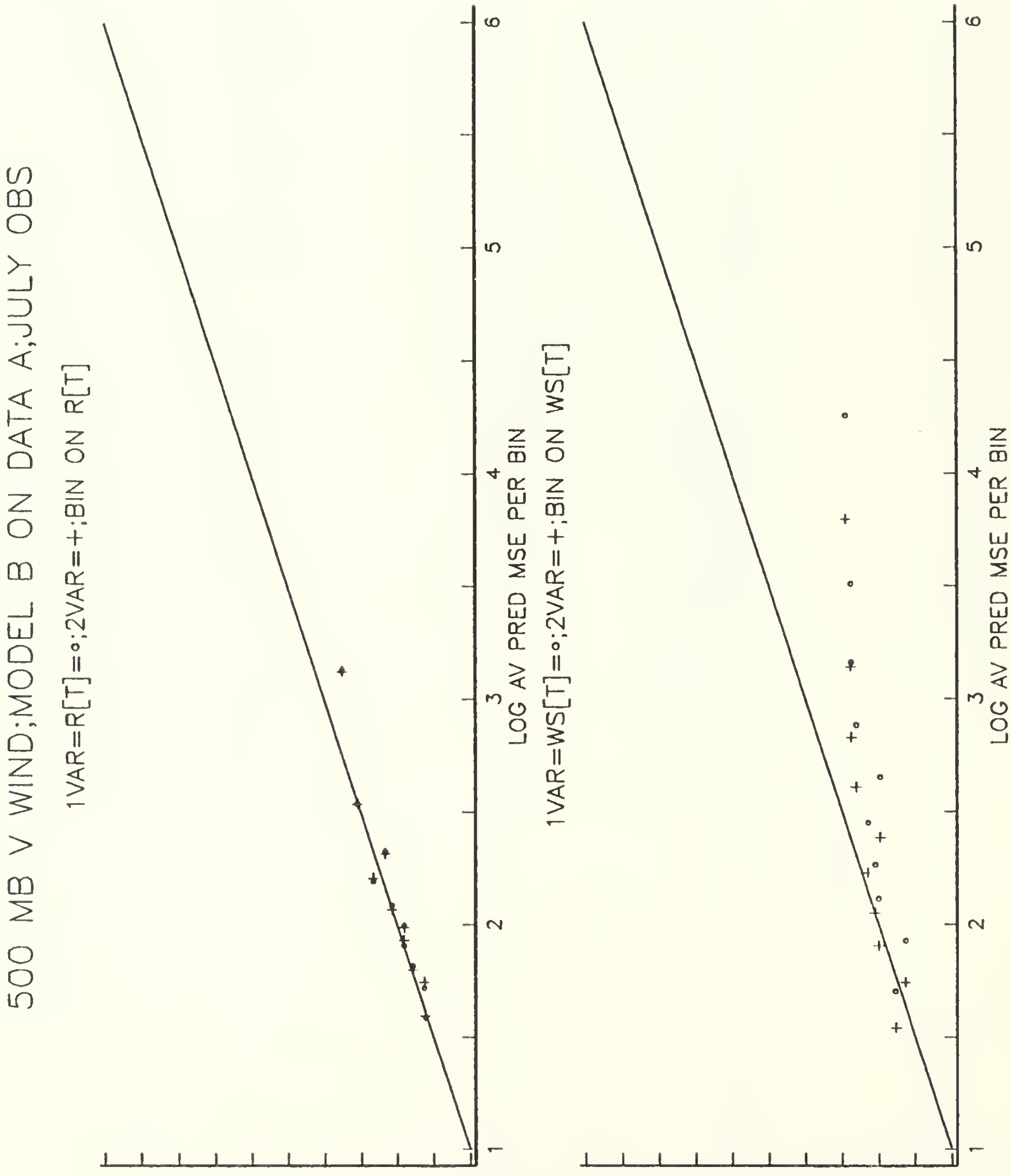

500 MB V WIND;MODEL B ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 15

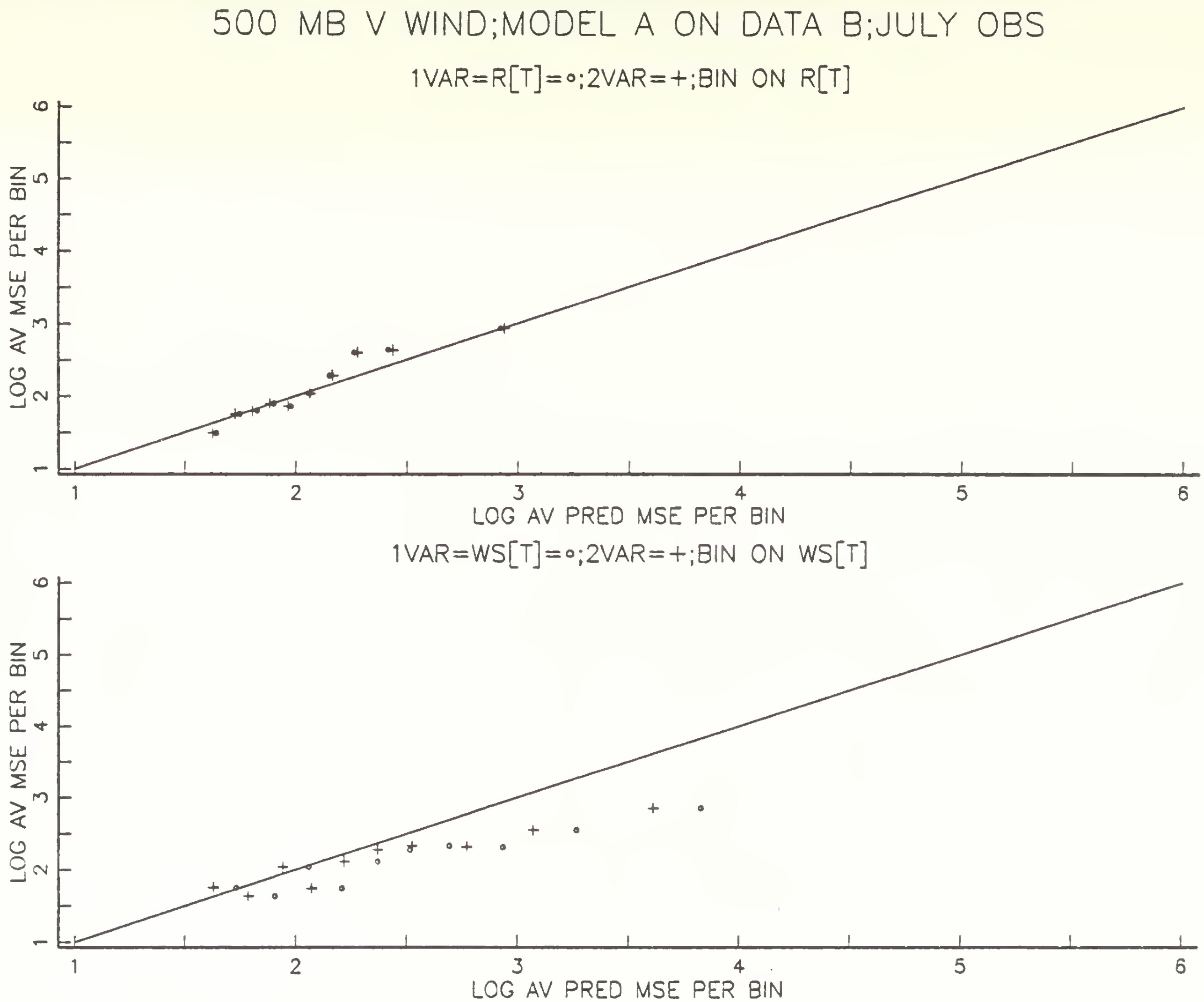

500 MB V WIND;MODEL A ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 16

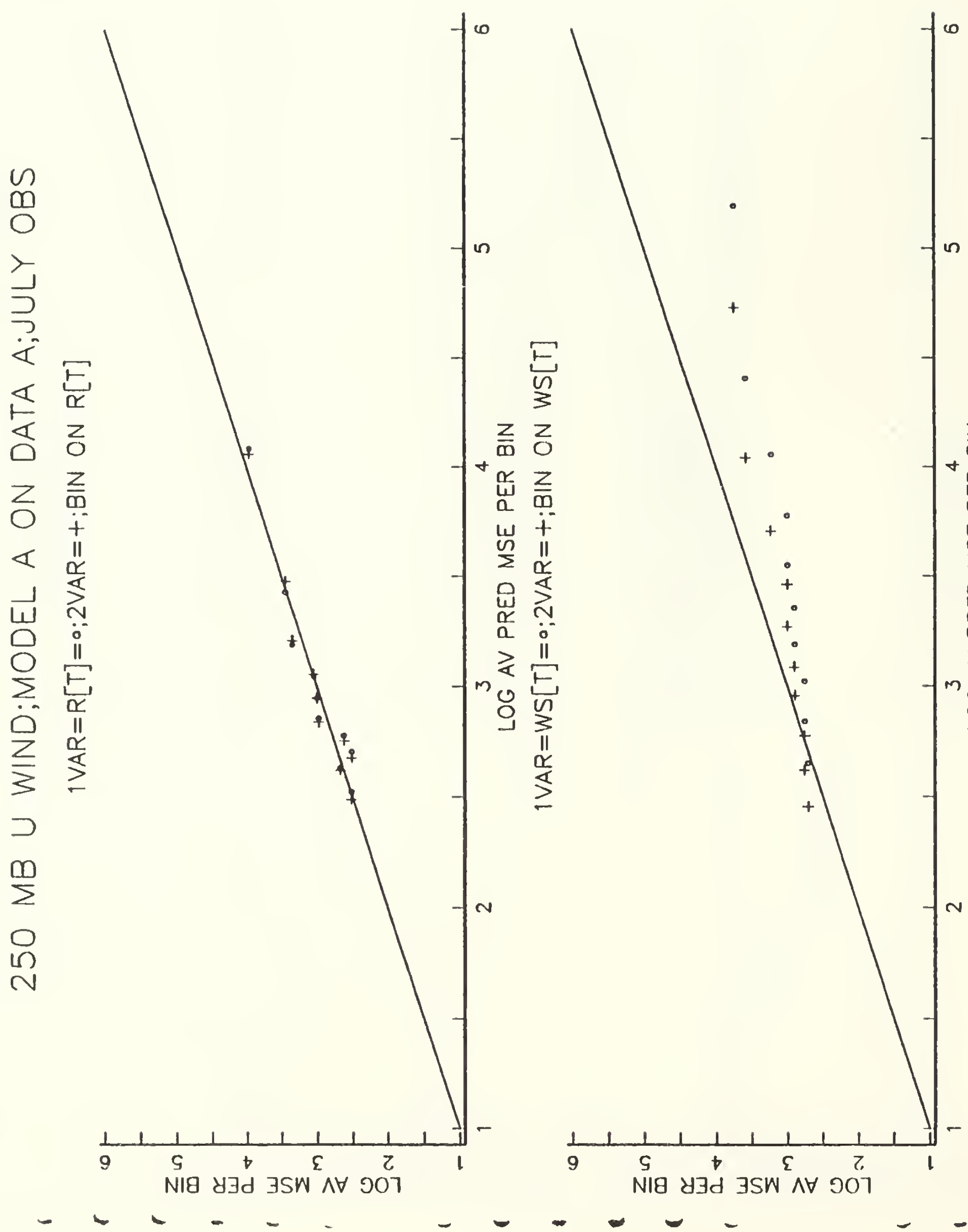

250 MB U WIND;MODEL A ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 17

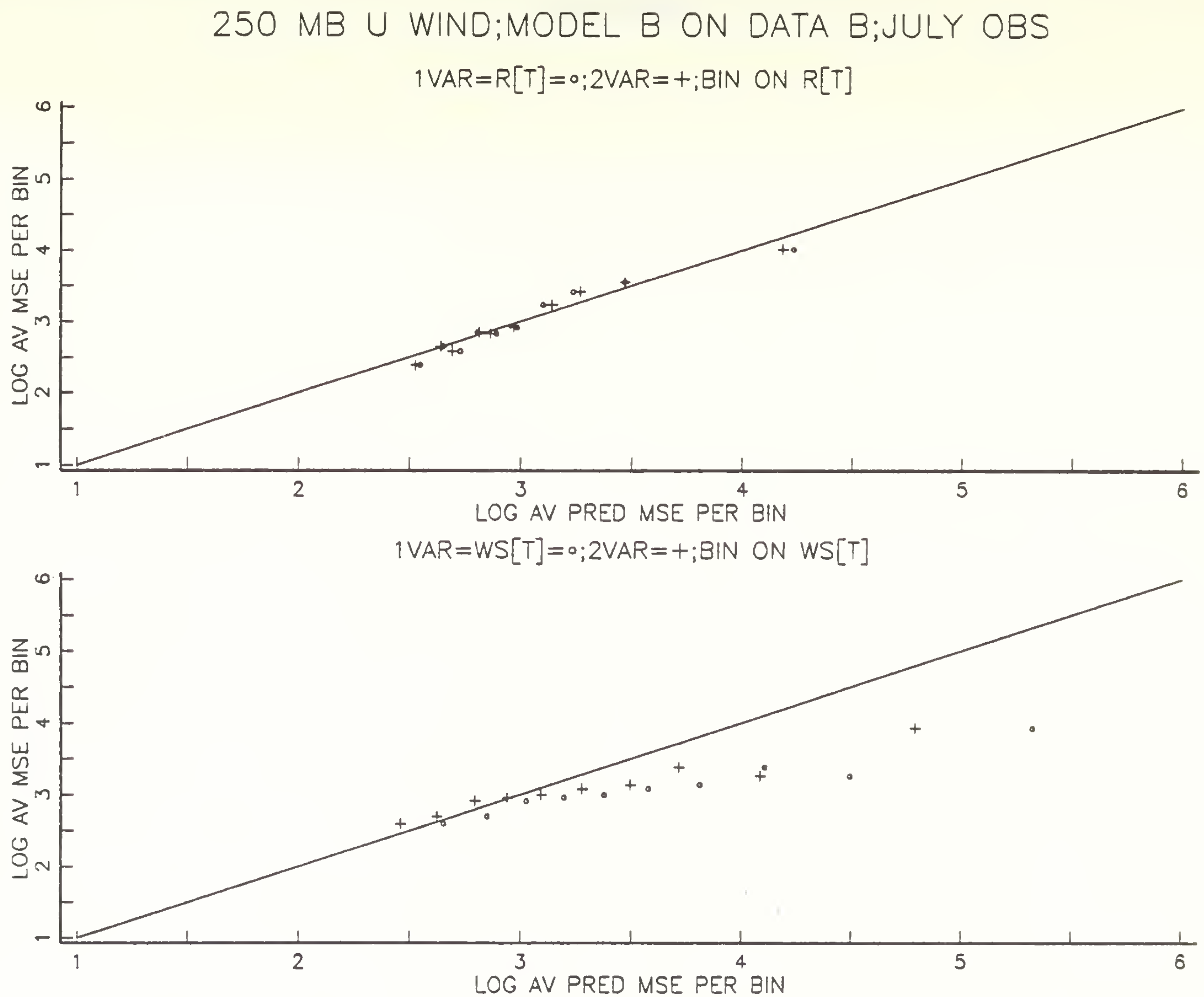

250 MB U WIND;MODEL B ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 18

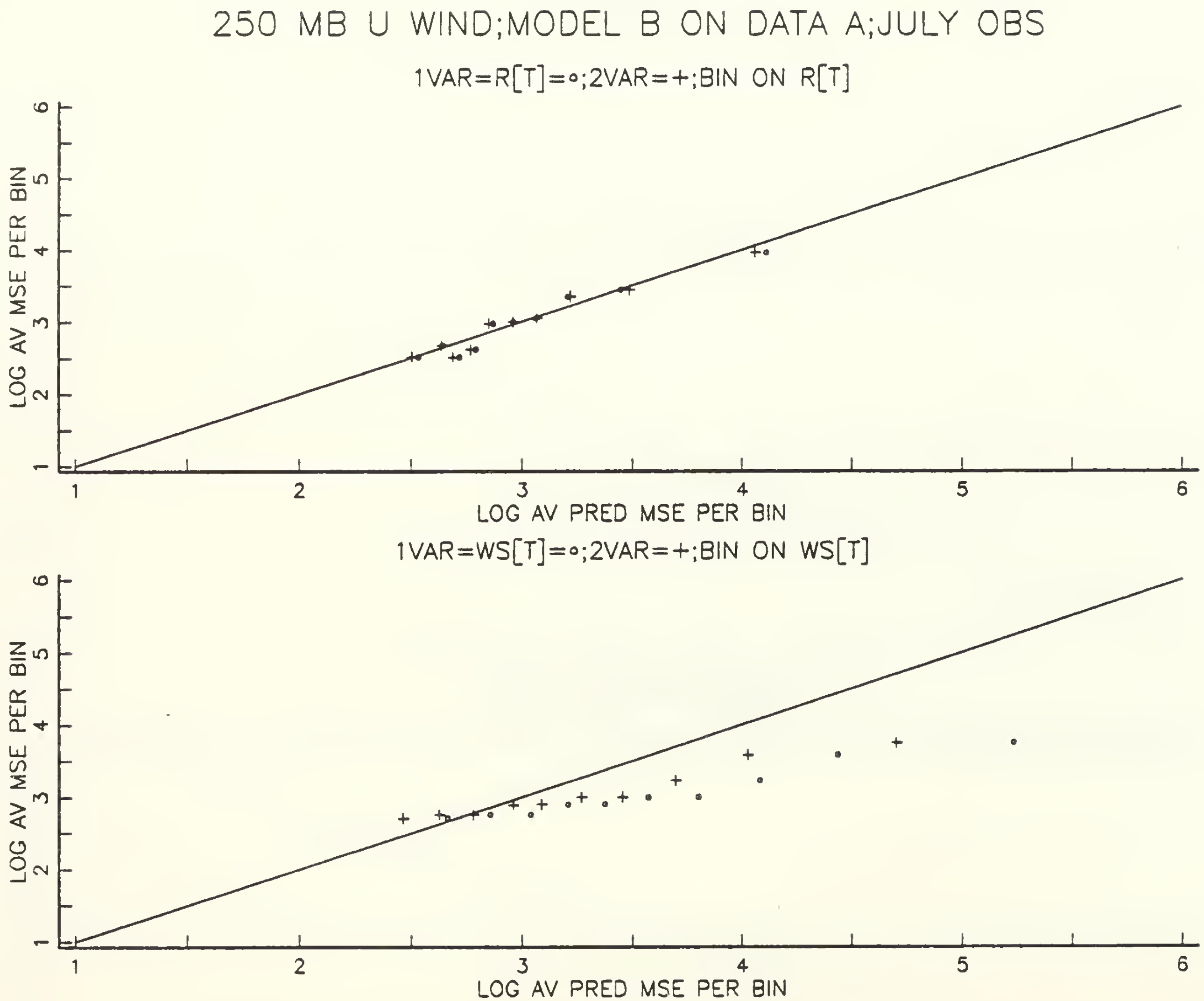

250 MB U WIND;MODEL B ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1
2
3
4
5
6

Figure 19

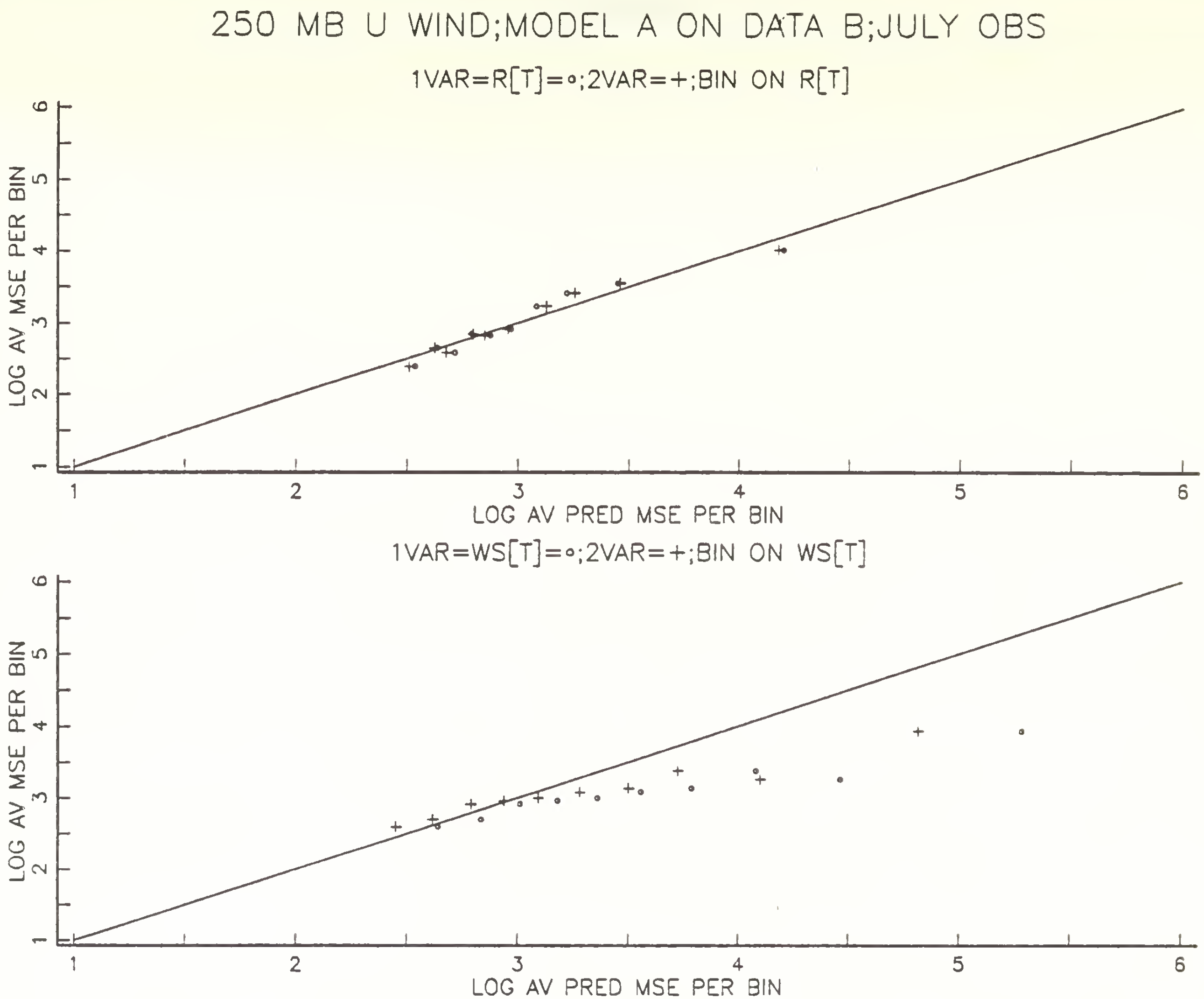

250 MB U WIND;MODEL A ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 20

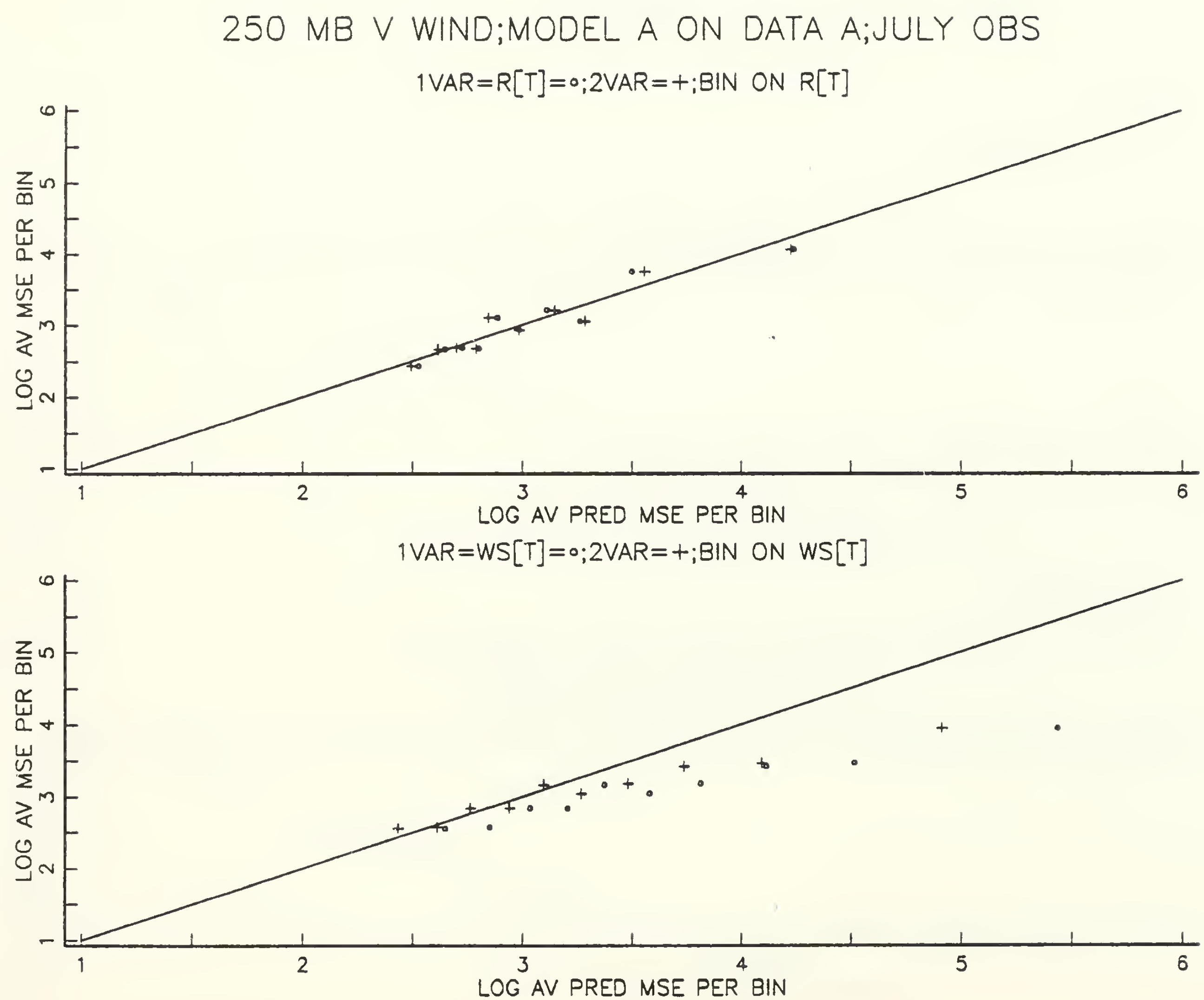

250 MB V WIND;MODEL A ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 21

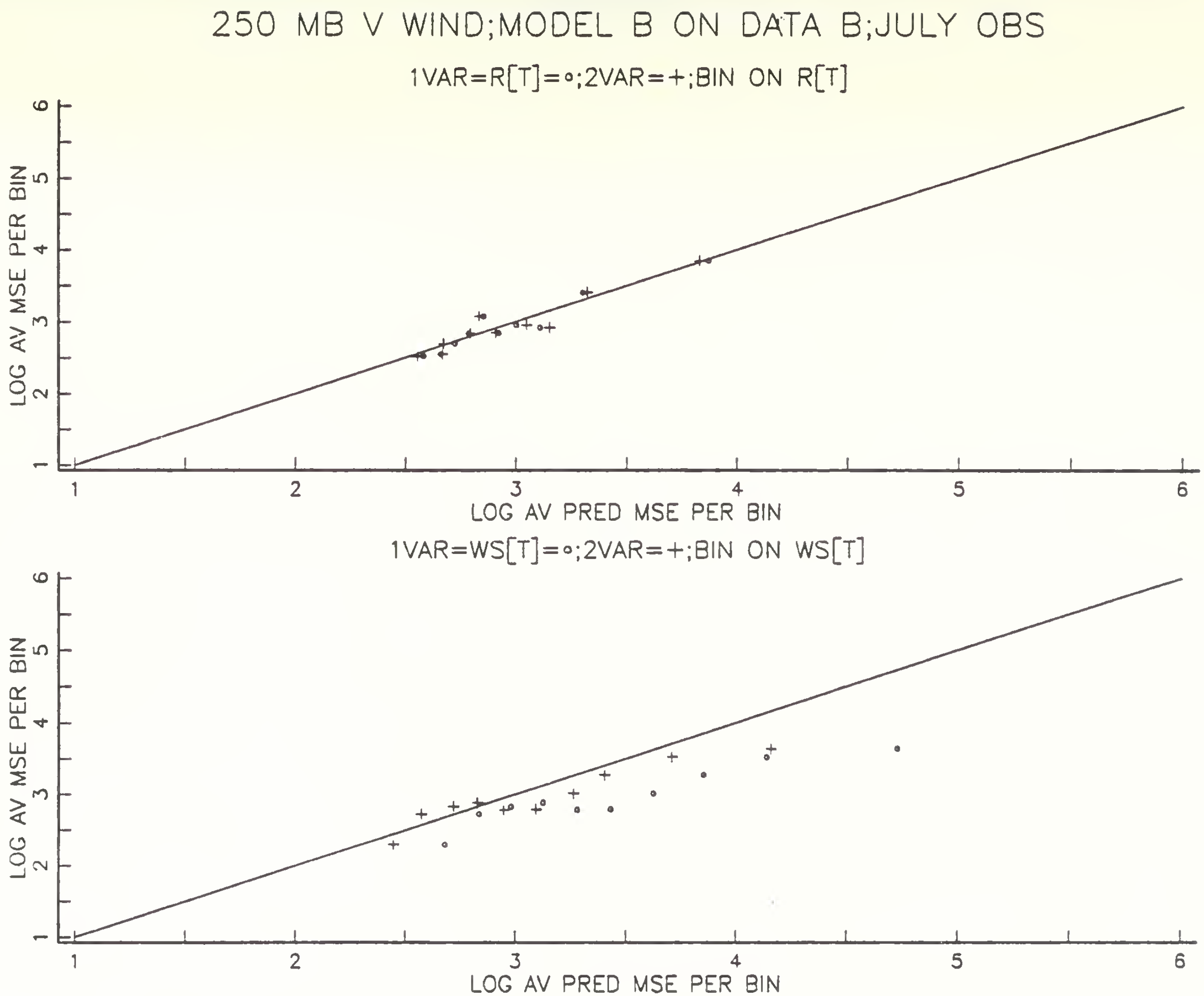

250 MB V WIND;MODEL B ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 22

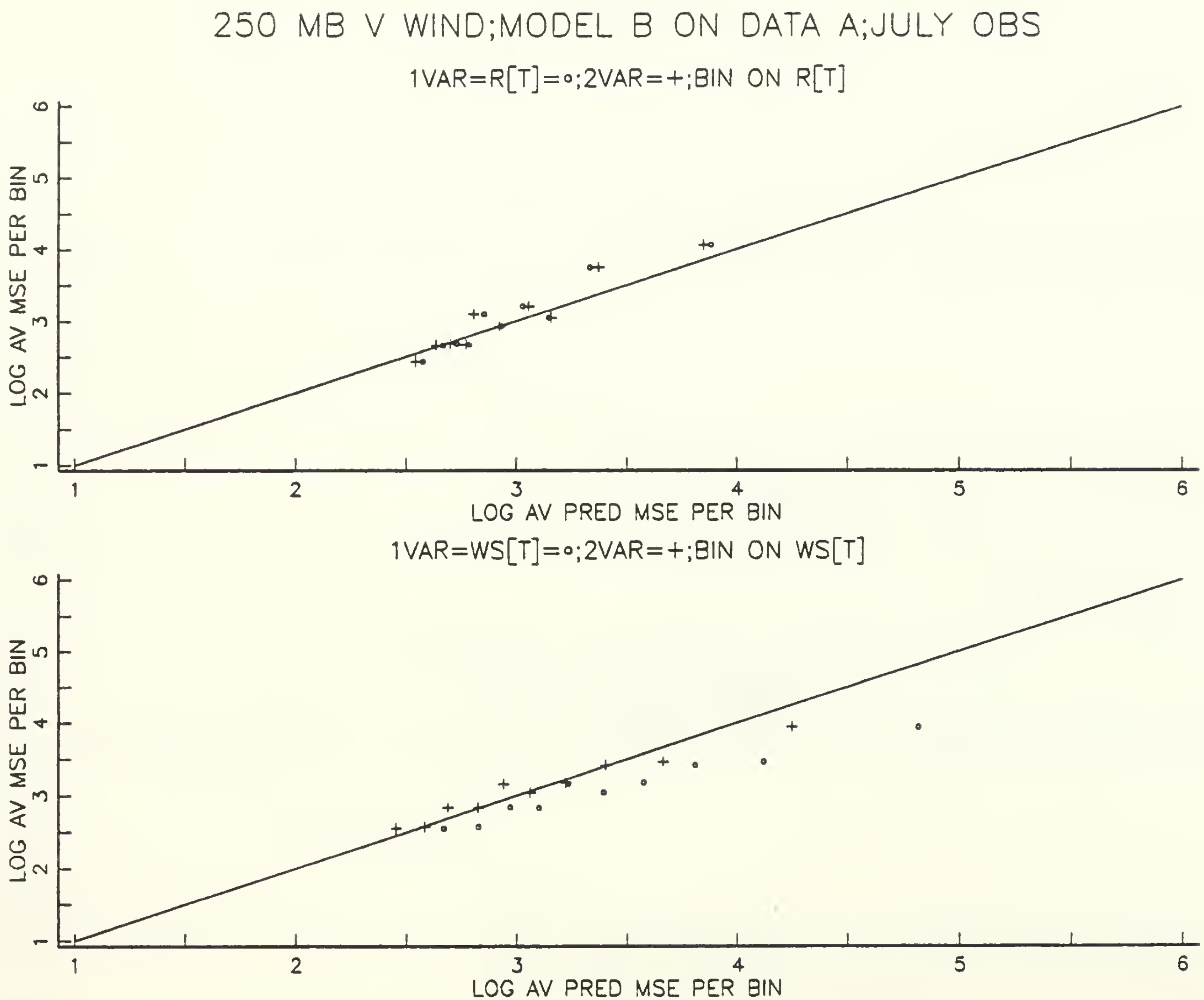

250 MB V WIND;MODEL B ON DATA A;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 23

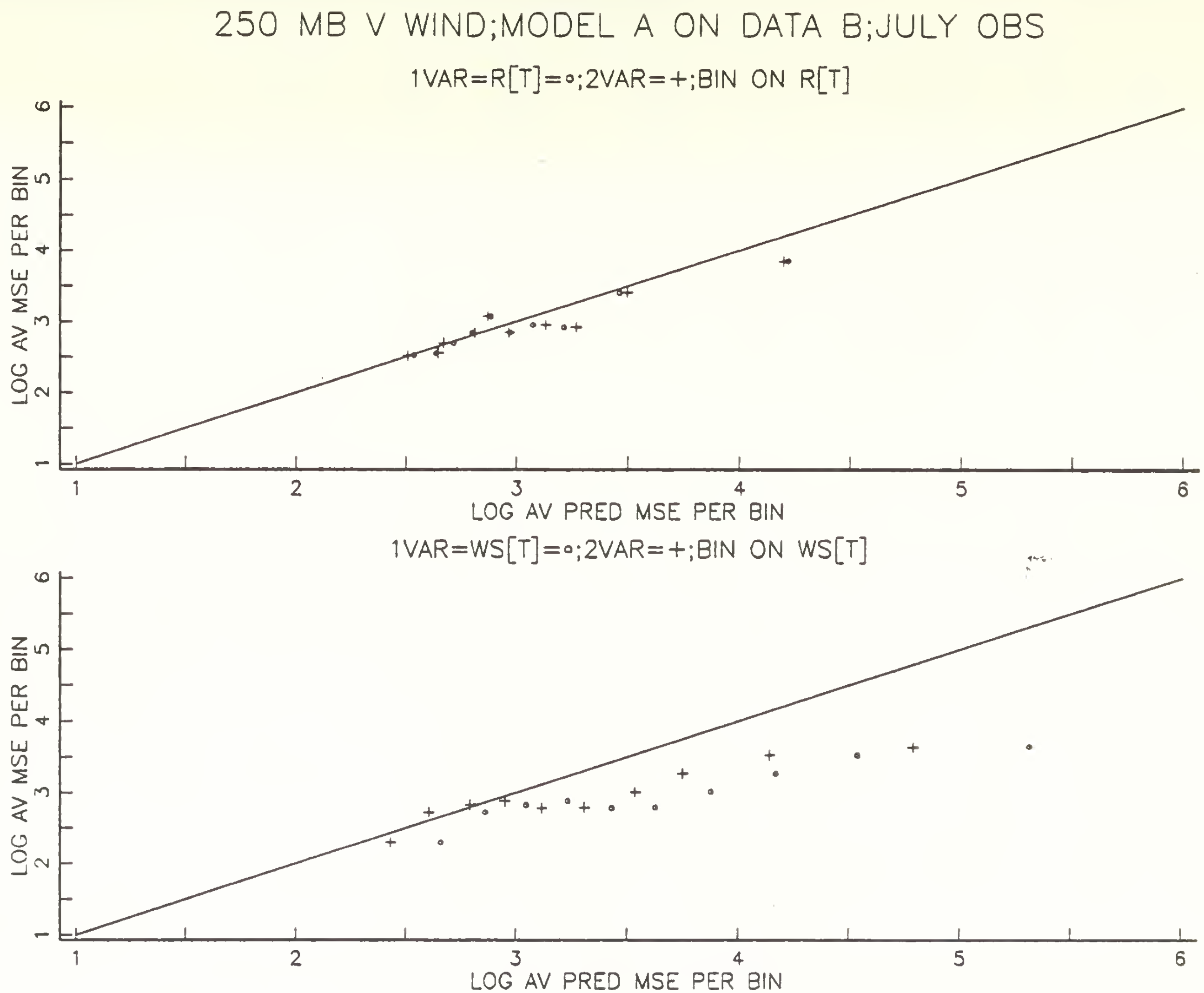

250 MB V WIND;MODEL A ON DATA B;JULY OBS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 24

REFERENCES

M. Aitken (1987). "Modeling variance heterogeneity in normal regression using GLIM." *Appl. Stat.*, **36**, pp. 332-339.

D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, Chapman and Hall, London, 1974.

M. Davidian and R. J. Carroll (1987) "Variance function estimation." *J. Am. Statis. Ass.*, **82**, pp. 1079-1091.

J. S. Goerss and P. A. Phoebus (1991a). *The Multivariate Optimum Interpolation Analysis of Meteorological Data at FNOC.* NOARL Report Number 31, Naval Oceanographic and Atmospheric Research Laboratory, Stennis Space Center, MS.

J. S. Goerss and P. A. Phoebus (1991b). "The Navy's operational atmospheric analysis." To appear in *Weather and Forecasting.*

T. F. Hogan and T. E. Rosmond (1991) "The description of the Navy operational global atmospheric prediction system's spectral forecast model." *Monthly Weather Review*, **119**, No. 8, pp. 1786-1815.

P. A. Jacobs and D. P. Gaver "Preliminary results from the analysis of wind component error," Naval Postgraduate School Techical Report NPSOR-91-029, September 1991.

P. McCullagh and J. A. Nelder. *Generalized Linear Regresion*, Chapman and Hall, New York, 1983.

J. A. Nelder and Y. Lee. (1992). "Likelihood, quasi-likelihood and pseudolikelihood: some comparisons." *J. R. Statis. Soc. B.*, **54**, No. 1, pp. 273-284.

APPENDIX A

A BOOTSTRAP CROSS-VALIDATION STUDY FOR JULY DATA

In this Appendix histograms are presented from a bootstrap cross-validation study of models for July using both observed wind covariates and first guess wind covariates. Figures 1A-6A present results for the observed wind covariates. Figures 7A-12A present results for the first guess wind covariates.

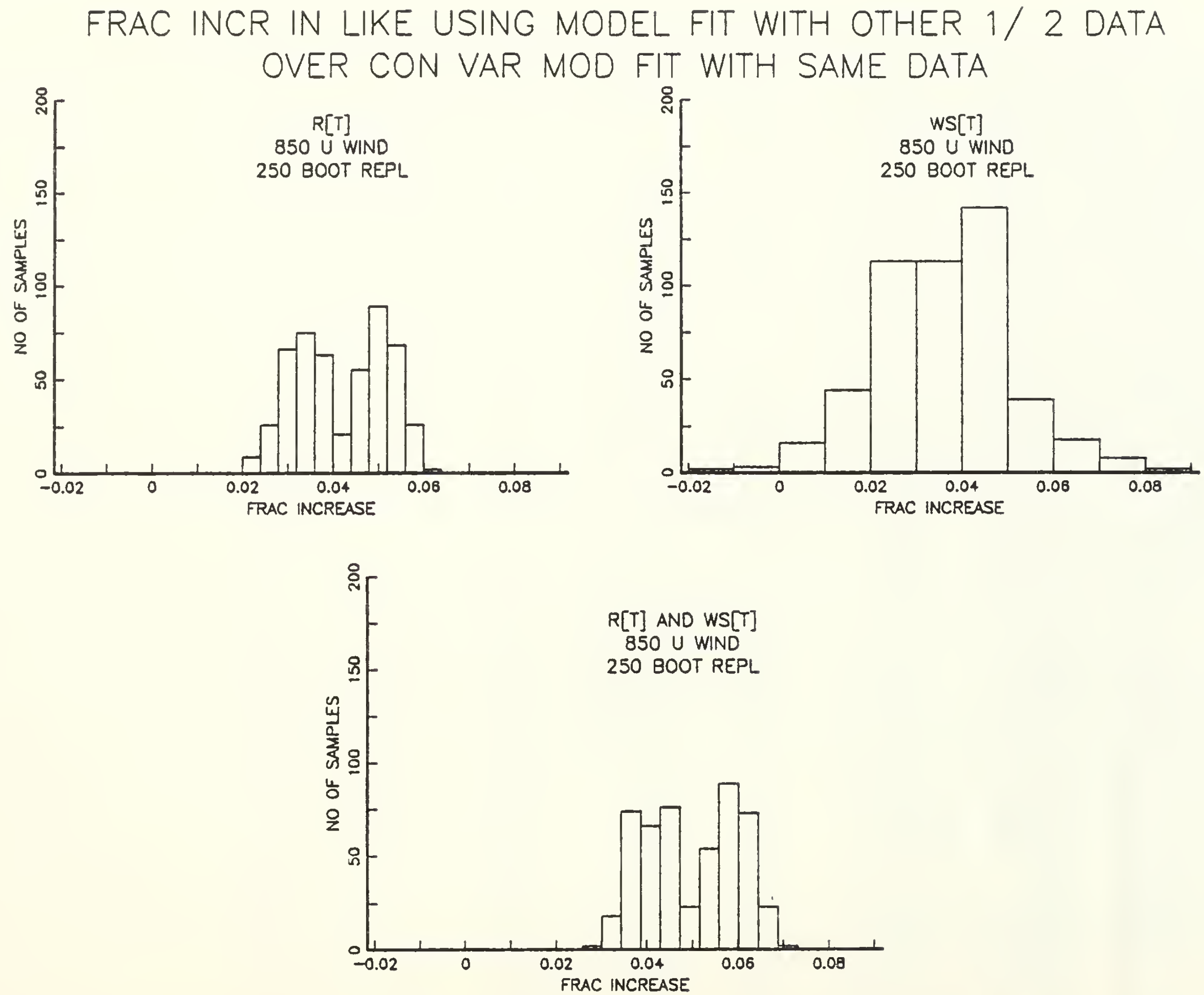
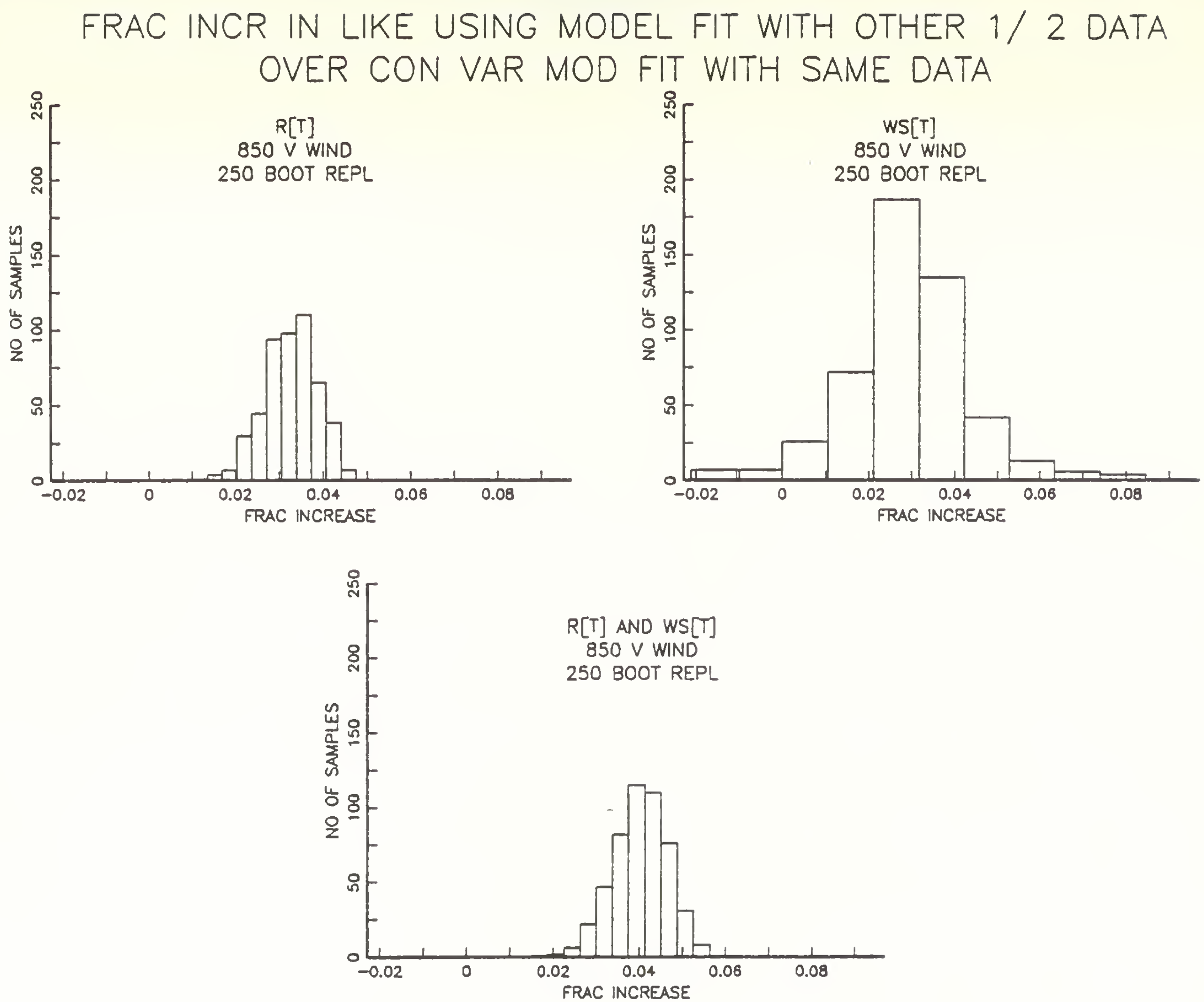

FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA
R[T]
850 U WIND
250 BOOT REPL
NO OF SAMPLES
FRAC INCREASE
WS[T]
850 U WIND
250 BOOT REPL
NO OF SAMPLES
FRAC INCREASE
R[T] AND WS[T]
850 U WIND
250 BOOT REPL
NO OF SAMPLES
FRAC INCREASE

Figure 1A



FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA
R[T]
850 V WIND
250 BOOT REPL
WS[T]
850 V WIND
250 BOOT REPL
R[T] AND WS[T]
850 V WIND
250 BOOT REPL
NO OF SAMPLES
FRAC INCREASE
0
50
100
150
200
250
−0.02
0
0.02
0.04
0.06
0.08

Figure 2A

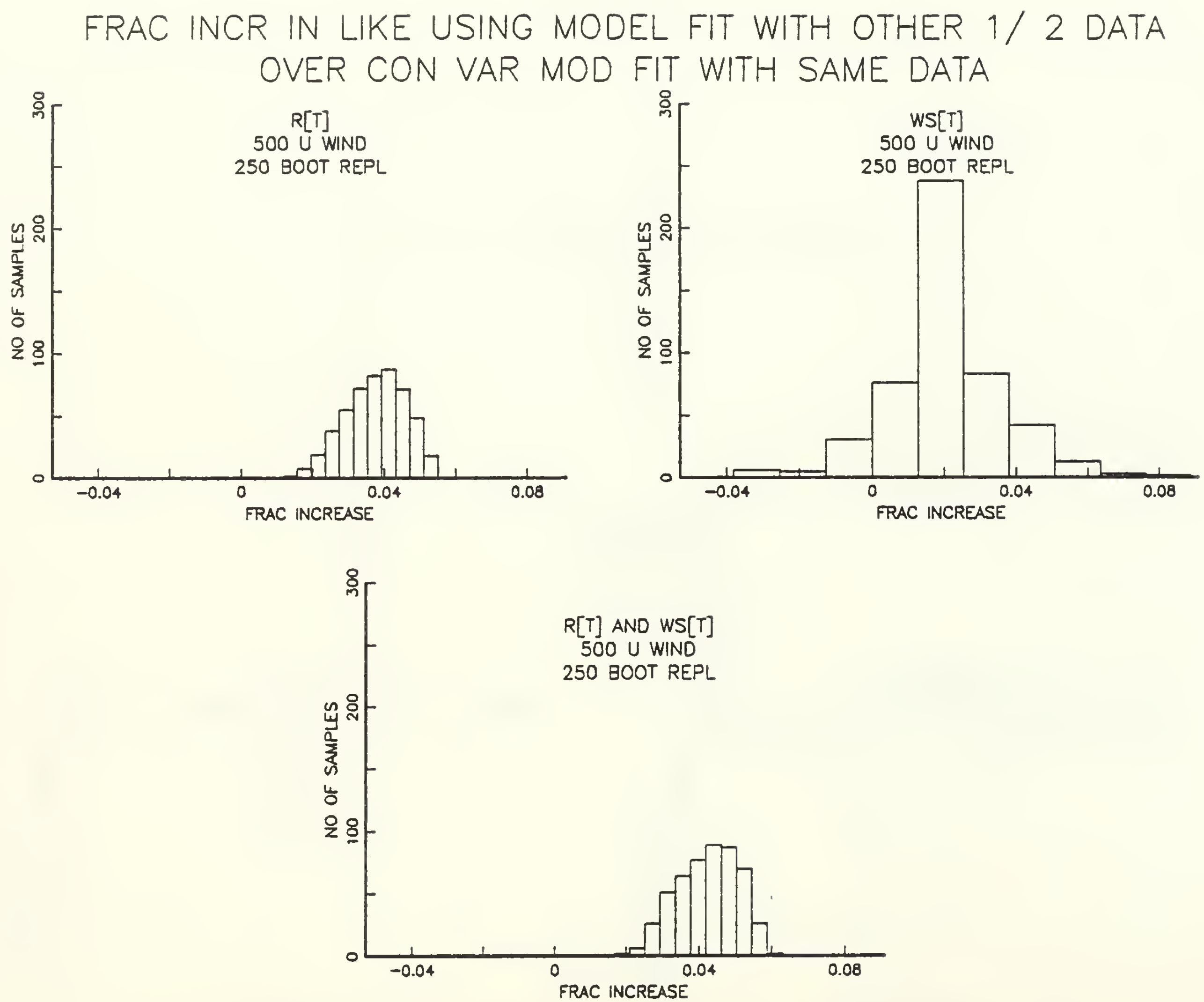

FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA
R[T]
500 U WIND
250 BOOT REPL
WS[T]
500 U WIND
250 BOOT REPL
R[T] AND WS[T]
500 U WIND
250 BOOT REPL
NO OF SAMPLES
FRAC INCREASE
0
100
200
300
-0.04
0
0.04
0.08

Figure 3A

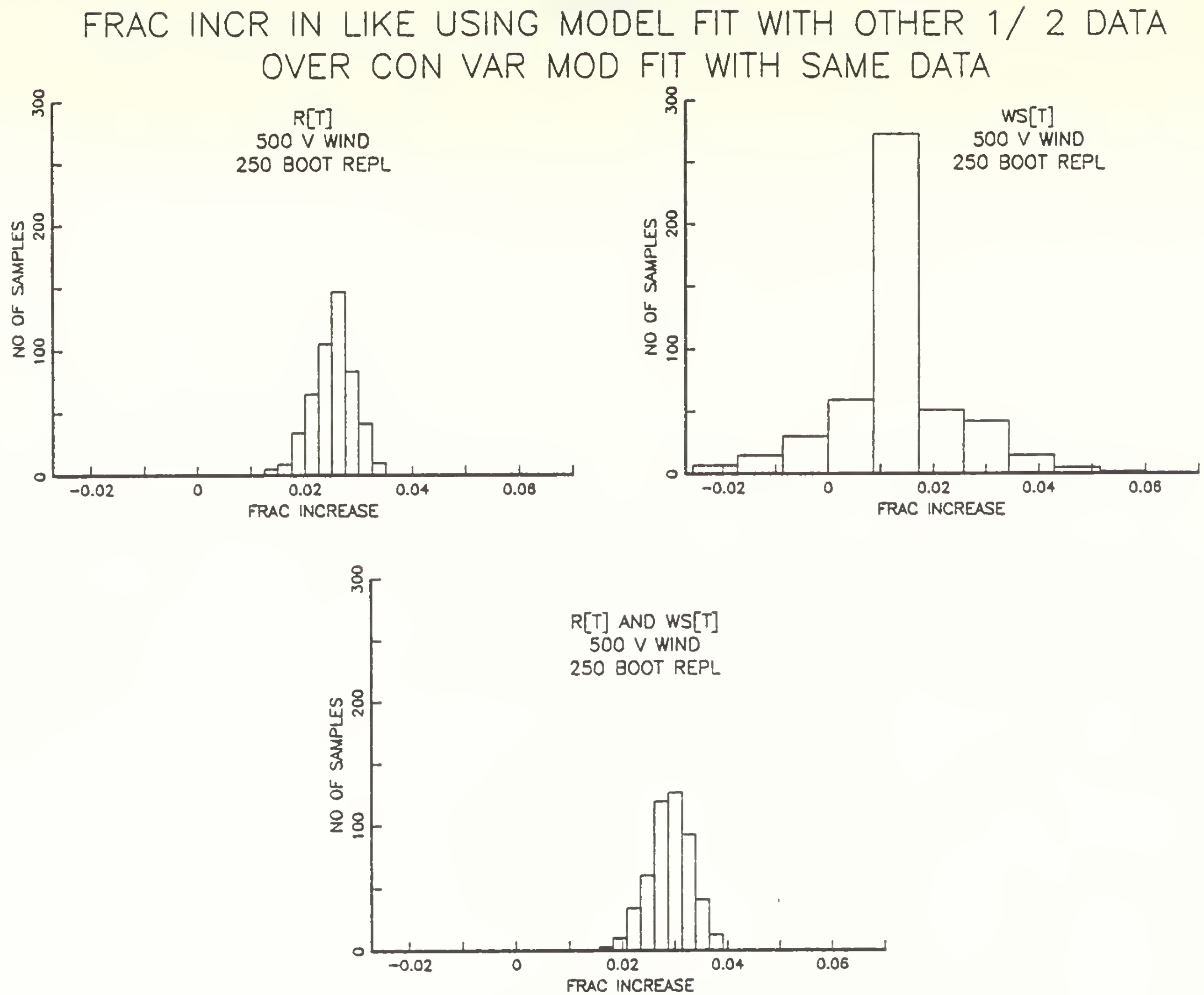

FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA
R[T]
500 V WIND
250 BOOT REPL
NO OF SAMPLES
0
100
200
300
-0.02
0
0.02
0.04
0.06
FRAC INCREASE
WS[T]
500 V WIND
250 BOOT REPL
NO OF SAMPLES
0
100
200
300
-0.02
0
0.02
0.04
0.06
FRAC INCREASE
R[T] AND WS[T]
500 V WIND
250 BOOT REPL
NO OF SAMPLES
0
100
200
300
-0.02
0
0.02
0.04
0.06
FRAC INCREASE

Figure 4A

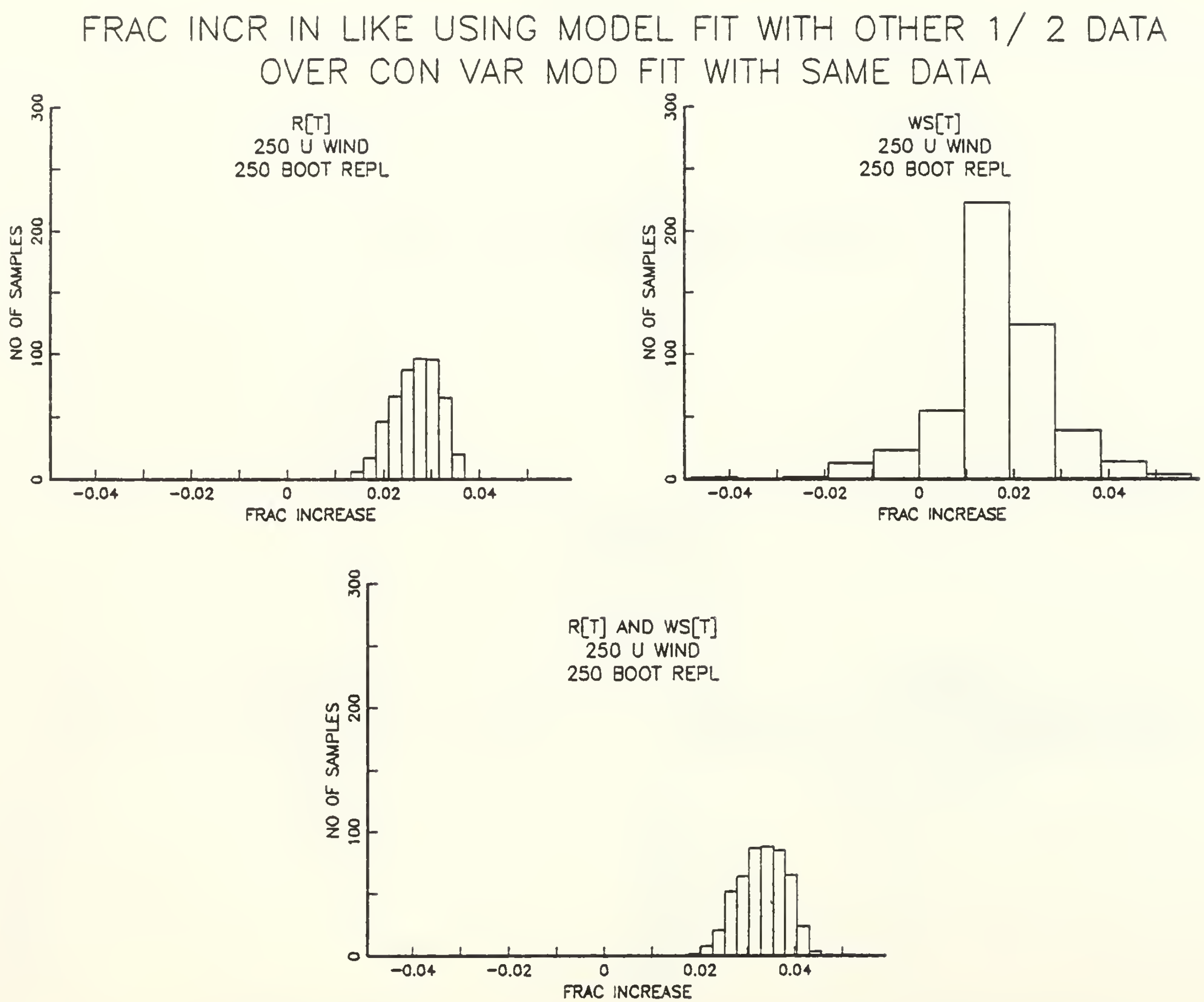

FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA
R[T]
250 U WIND
250 BOOT REPL
NO OF SAMPLES
FRAC INCREASE
WS[T]
250 U WIND
250 BOOT REPL
NO OF SAMPLES
FRAC INCREASE
R[T] AND WS[T]
250 U WIND
250 BOOT REPL
NO OF SAMPLES
FRAC INCREASE

Figure 5A

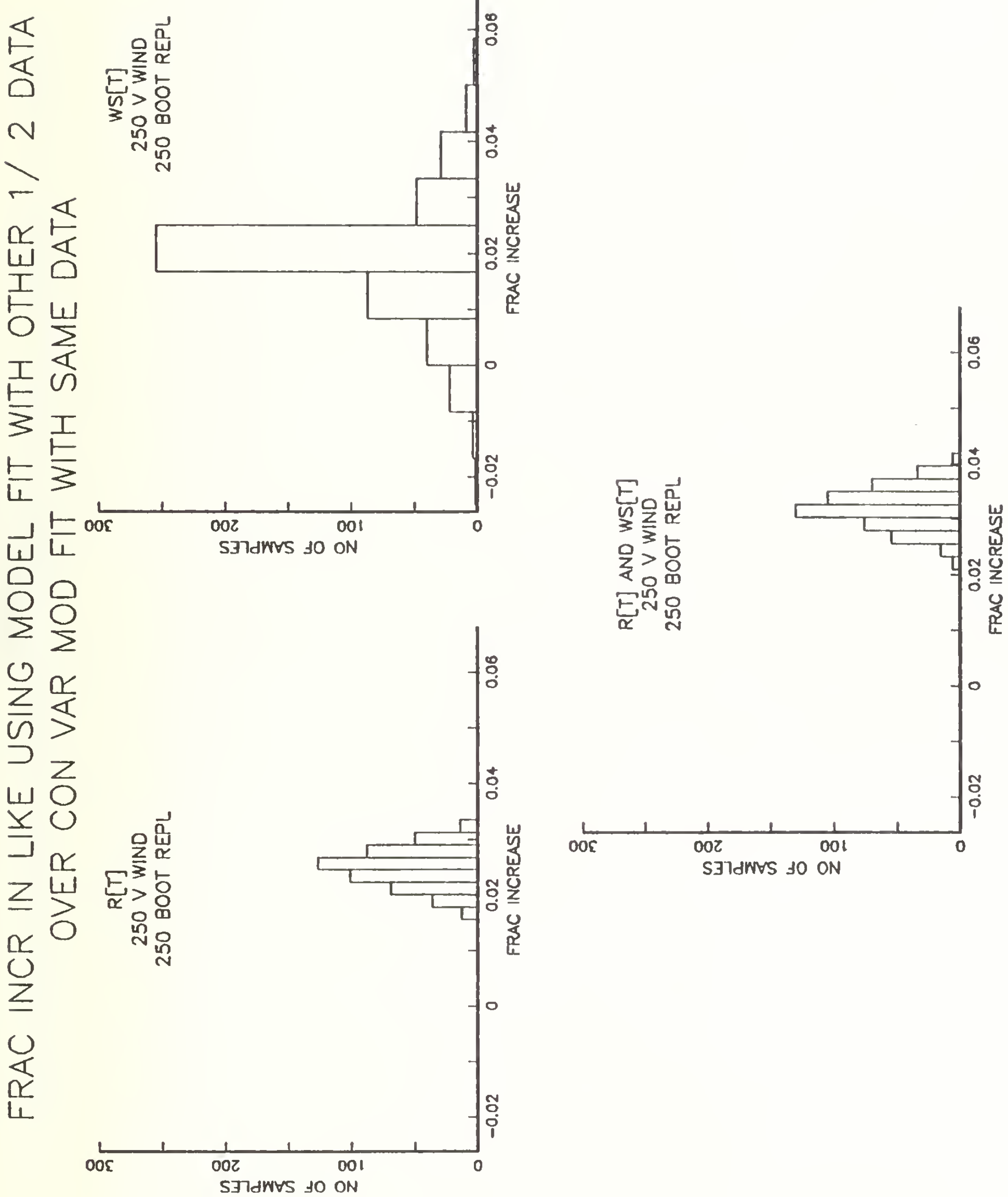

FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA
R[T]
250 V WIND
250 BOOT REPL
WS[T]
250 V WIND
250 BOOT REPL
R[T] AND WS[T]
250 V WIND
250 BOOT REPL
NO OF SAMPLES
FRAC INCREASE
0
100
200
300
−0.02
0
0.02
0.04
0.06

Figure 6A

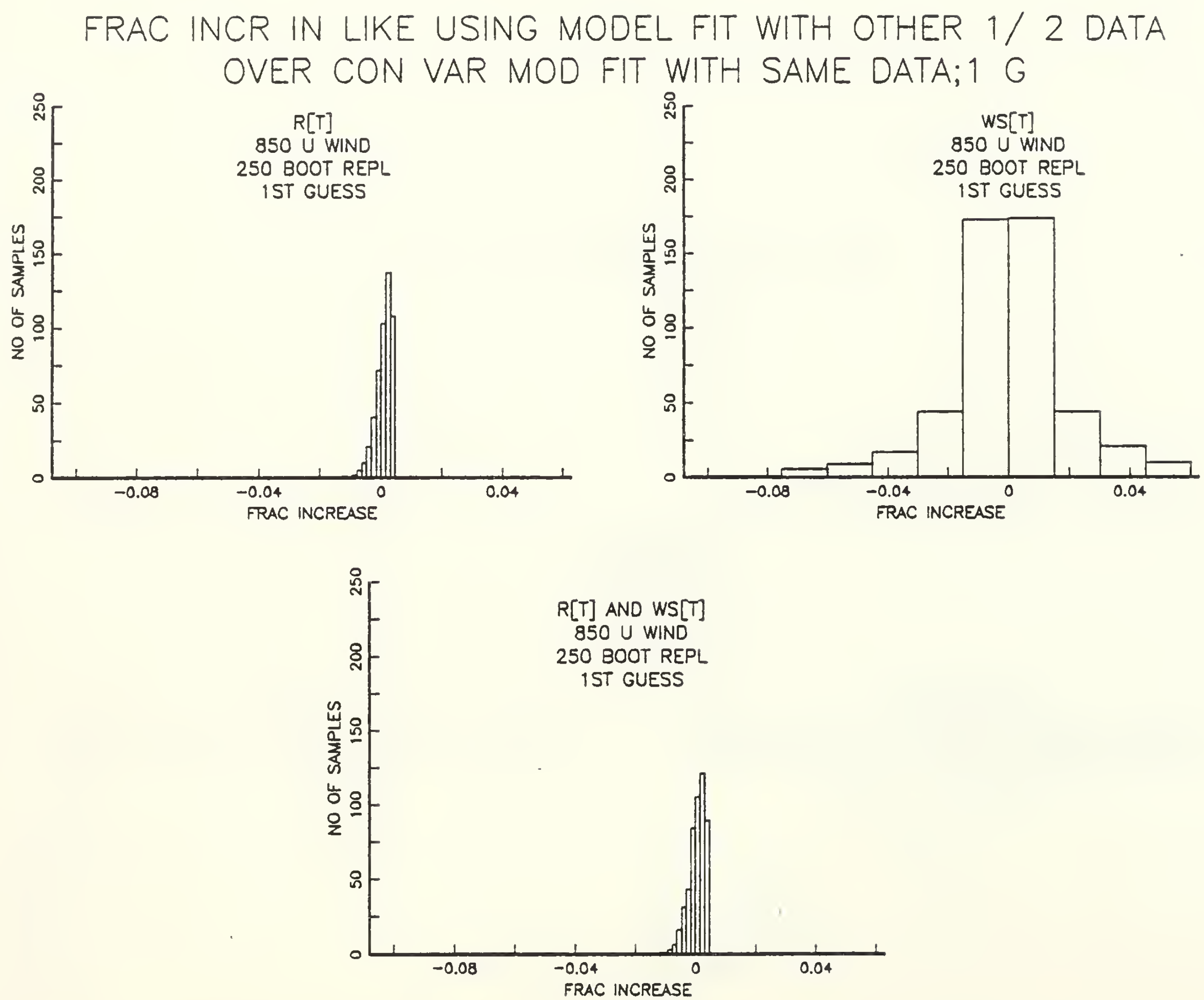

FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA;1 G
R[T]
850 U WIND
250 BOOT REPL
1ST GUESS
NO OF SAMPLES
0
50
100
150
200
250
-0.08
-0.04
0
0.04
FRAC INCREASE
WS[T]
850 U WIND
250 BOOT REPL
1ST GUESS
R[T] AND WS[T]
850 U WIND
250 BOOT REPL
1ST GUESS

Figure 7A

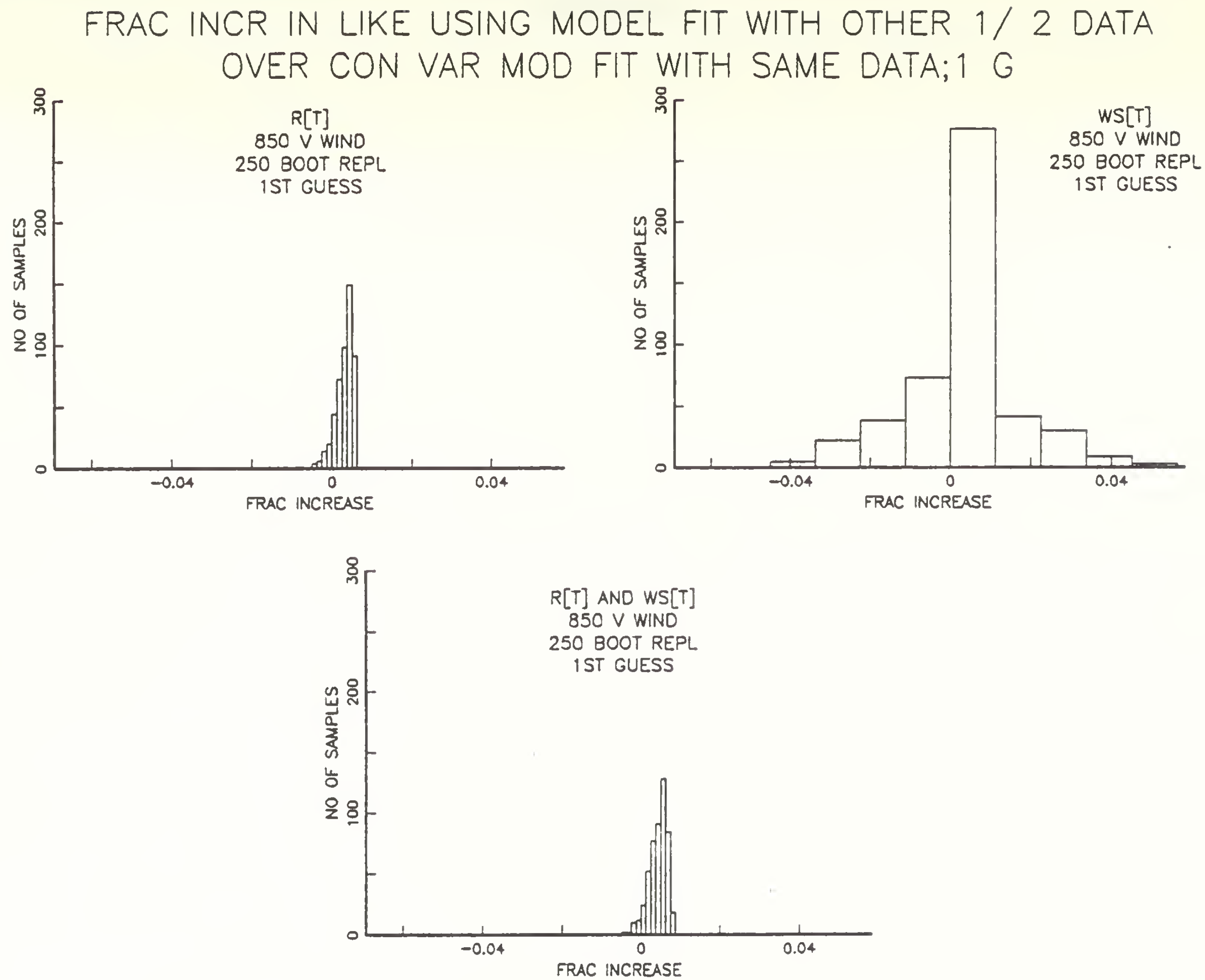

FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA;1 G
R[T]
850 V WIND
250 BOOT REPL
1ST GUESS
NO OF SAMPLES
FRAC INCREASE
WS[T]
850 V WIND
250 BOOT REPL
1ST GUESS
NO OF SAMPLES
FRAC INCREASE
R[T] AND WS[T]
850 V WIND
250 BOOT REPL
1ST GUESS
NO OF SAMPLES
FRAC INCREASE

Figure 8A

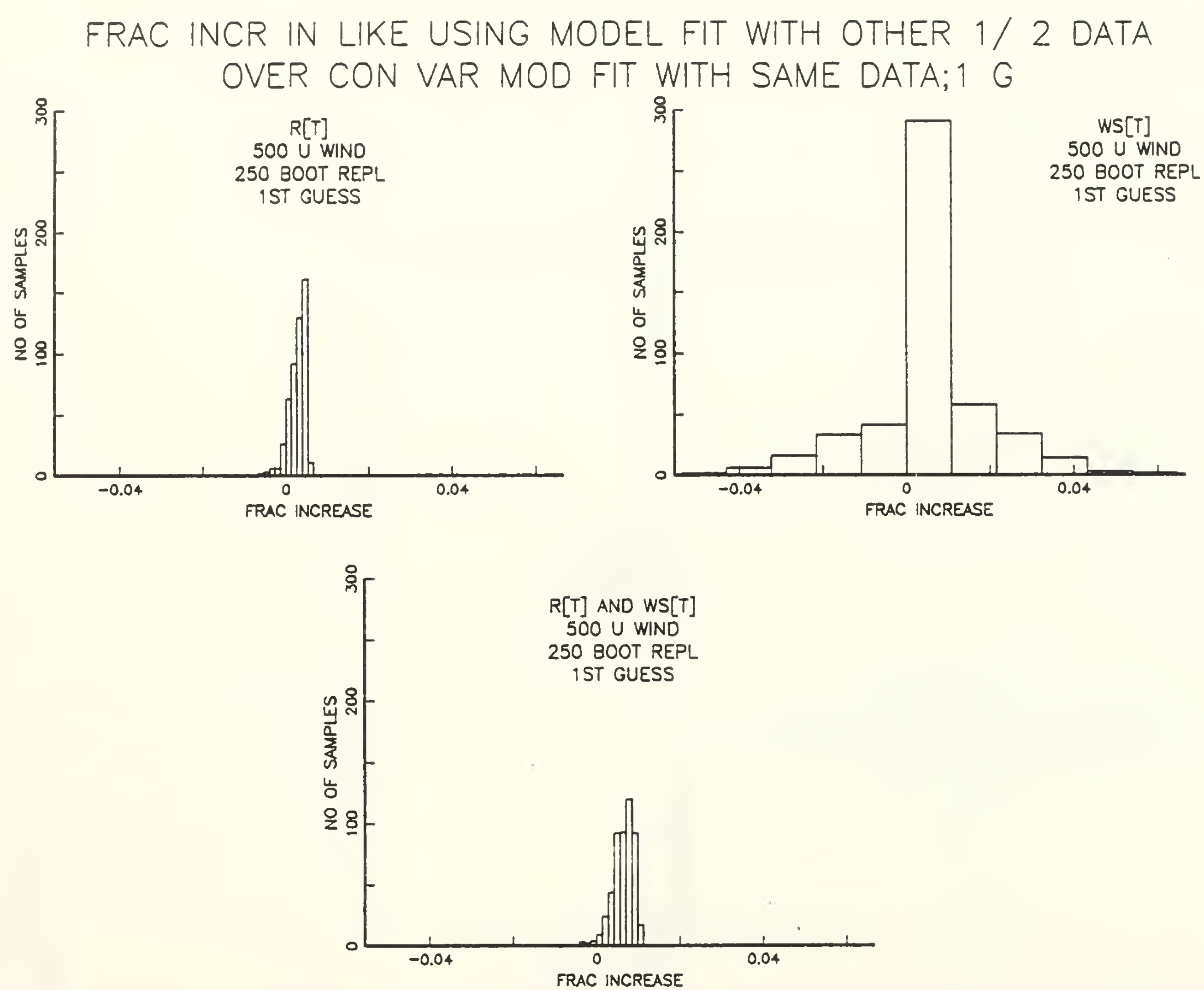

FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA;1 G
R[T]
500 U WIND
250 BOOT REPL
1ST GUESS
WS[T]
500 U WIND
250 BOOT REPL
1ST GUESS
R[T] AND WS[T]
500 U WIND
250 BOOT REPL
1ST GUESS
NO OF SAMPLES
0
100
200
300
FRAC INCREASE
-0.04
0
0.04

Figure 9A

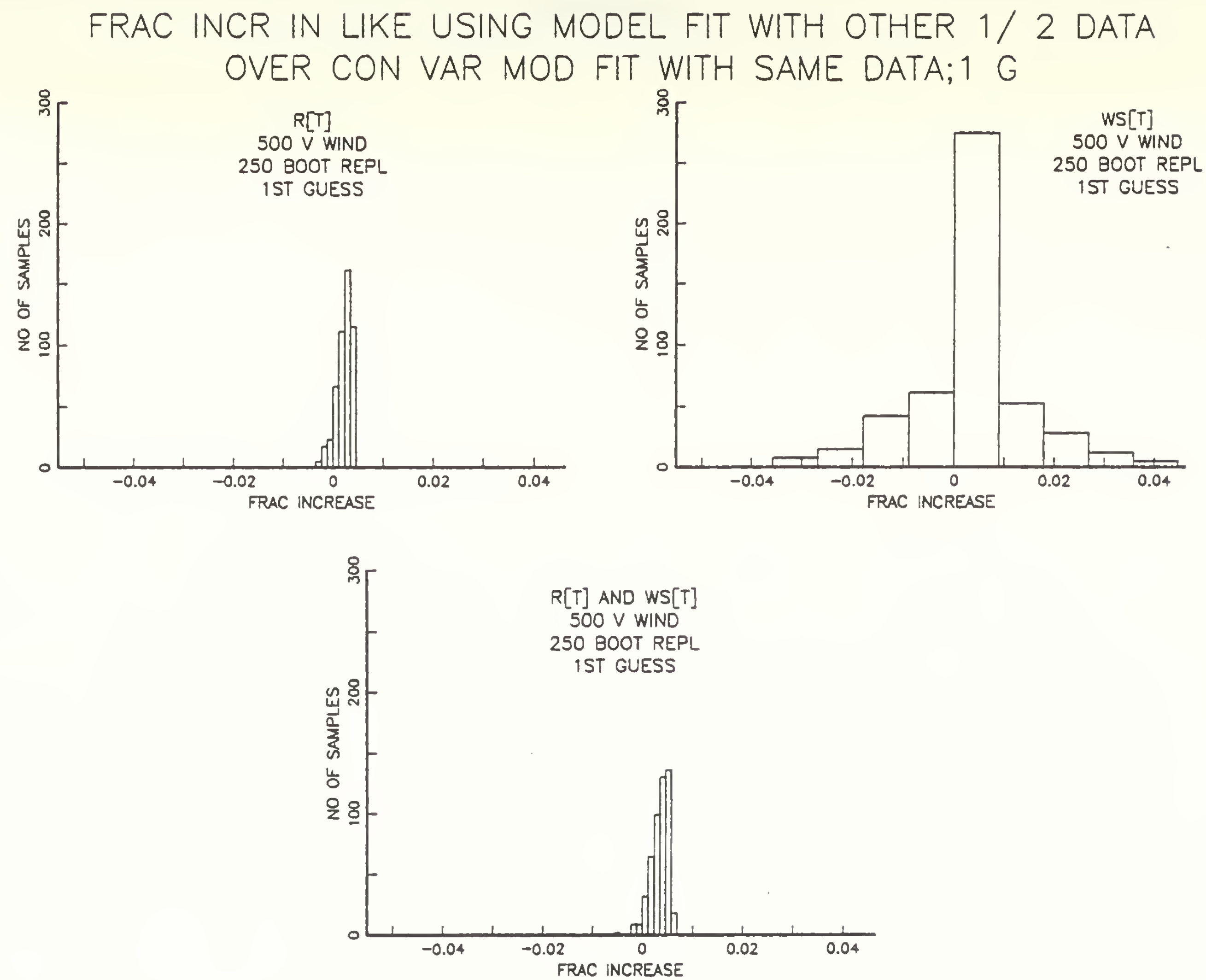

FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA;1 G
R[T]
500 V WIND
250 BOOT REPL
1ST GUESS
NO OF SAMPLES
FRAC INCREASE
WS[T]
500 V WIND
250 BOOT REPL
1ST GUESS
NO OF SAMPLES
FRAC INCREASE
R[T] AND WS[T]
500 V WIND
250 BOOT REPL
1ST GUESS
NO OF SAMPLES
FRAC INCREASE

Figure 10A

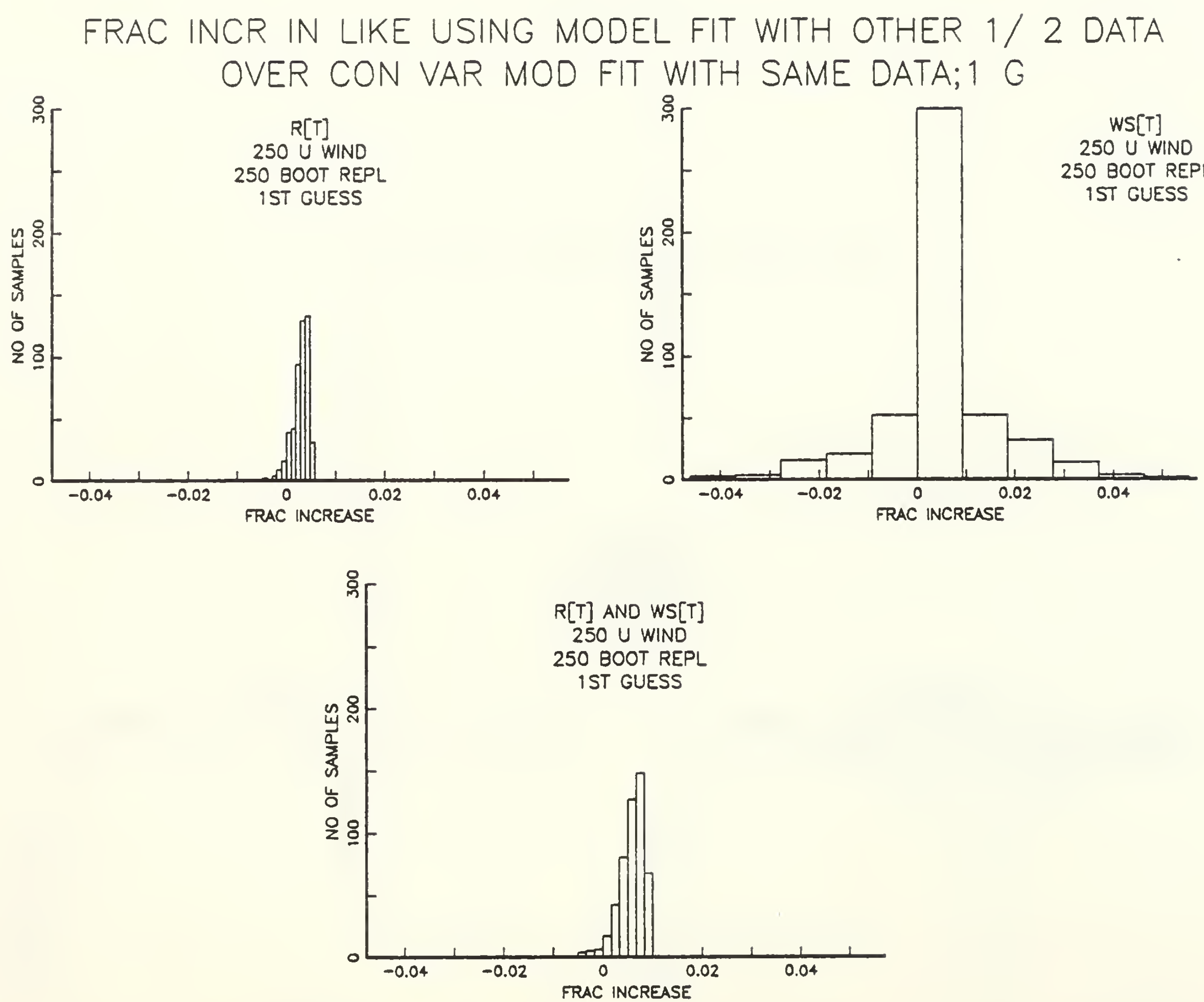
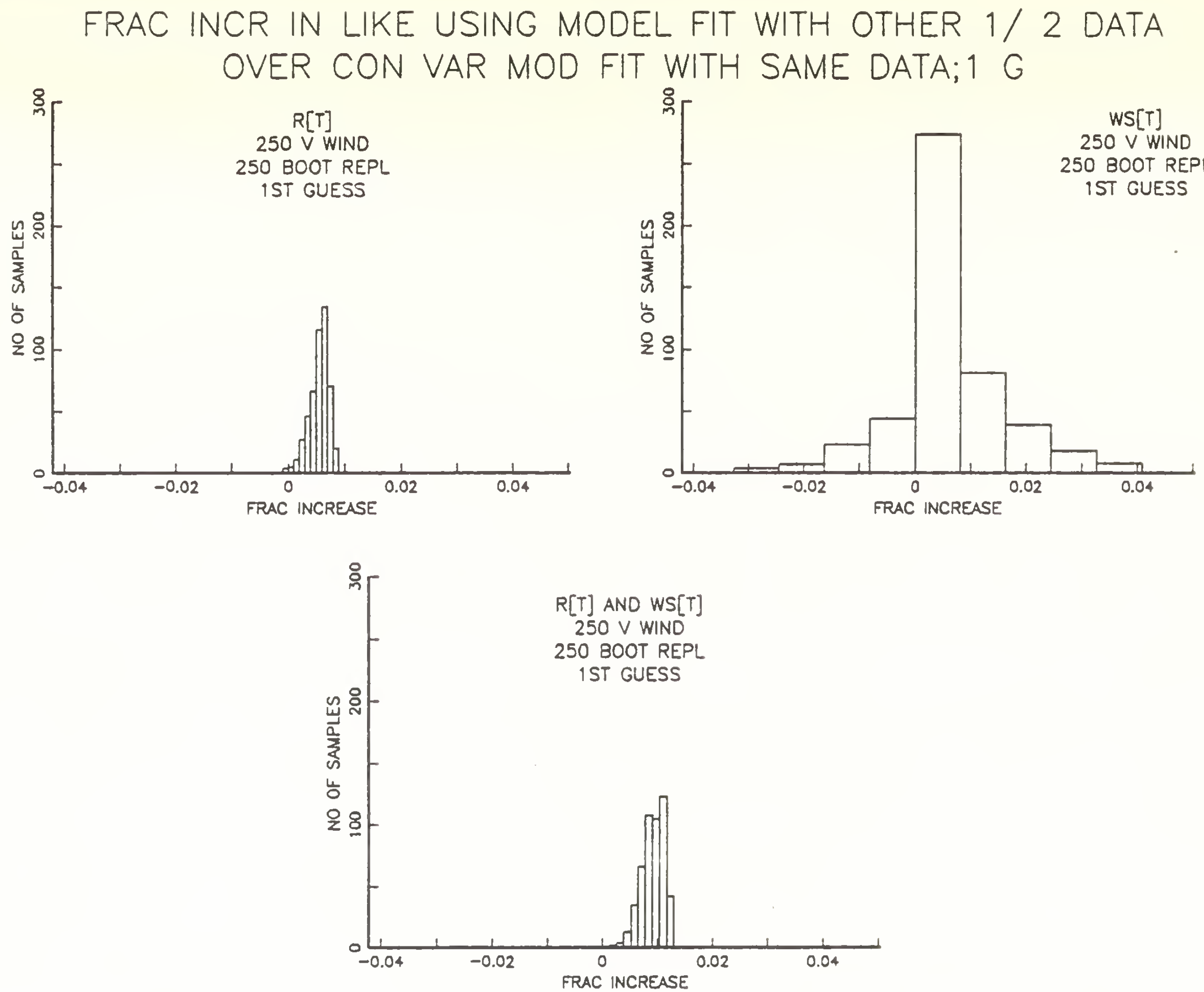

FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA;1 G
R[T]
250 U WIND
250 BOOT REPL
1ST GUESS
WS[T]
250 U WIND
250 BOOT REPL
1ST GUESS
R[T] AND WS[T]
250 U WIND
250 BOOT REPL
1ST GUESS
NO OF SAMPLES
FRAC INCREASE
0
100
200
300
-0.04
-0.02
0.02
0.04

Figure 11A



FRAC INCR IN LIKE USING MODEL FIT WITH OTHER 1/ 2 DATA
OVER CON VAR MOD FIT WITH SAME DATA;1 G
R[T]
250 V WIND
250 BOOT REPL
1ST GUESS
WS[T]
250 V WIND
250 BOOT REPL
1ST GUESS
R[T] AND WS[T]
250 V WIND
250 BOOT REPL
1ST GUESS
NO OF SAMPLES
FRAC INCREASE
0
100
200
300
-0.04
-0.02
0
0.02
0.04

Figure 12A

APPENDIX B

A GRAPHICAL ASSESSMENT OF GOODNESS OF FIT AND CROSS-VALIDATION OF MODELS OF JULY WIND COMPONENT MEAN SQUARE ERROR USING FIRST-GUESS WIND COVARIATES

In this appendix we present figures assessing goodness of fit and cross-validation of the normal models (1)–(3) with first-guess wind covariates fit to July data. As in subsection (3.2) the data is randomly divided into two sets called DA and DB without regard to the values of the data; these sets are the same as those in that section.

The maximum likelihood parameter estimates for each model (1)–(3) are obtained for each set DA and DB and appear in Table 4. The estimated variances $\sigma_1^2(1,t)$, $\sigma_1^2(2,t)$, $\sigma_2^2(t)$ are computed for the parameters estimated from DA and DB using (1)-(3) for each data point in DA and DB.

To assess models (1) and (3) the data $(y(t), r(t), s(t))$ are binned into 10 bins based on ordering the values of $r(t)$ from smallest to largest. The data in the first bin correspond to the smaller values of $r(t)$; the data in the 10^{th} bin correspond to the larger values of $r(t)$. Each bin contains about $\frac{1}{10}^{th}$ of the data with the 10^{th} bin containing a few more data. The averages of the estimated variances for models (1) and (3) are computed for each bin. The average $y(t)^2$ is also computed for each bin.

To assess models (2) and (3) the same procedure is used but the binning is based on values of $s(t)$.

Figures 1B-24B present graphs of the log[average $y(t)^2$] in each bin versus log[average estimated variance] in each bin for models (1) and (3) and models

(2) and (3). Figures 1B, 5B, 9B, 13B, 17B, 21B (respectively 2B, 6B, 10B, 14B, 18B 22B) show the logarithm of the average of the $y(t)^2$ values of DA (respectively DB) versus the logarithm of the average of the estimated variances for each bin using the estimated parameters from DA (respectively DB). If a model were perfect, a point should be close to the 45° line shown. These figures assess goodness of fit.

Figures 3B, 7B, 11B, 15B, 19B, 23B (respectively 4B, 8B, 12B, 16B, 20B, 24B) present graphs of log average $y(t)^2$ of DA (respectively DB) versus log average estimated variances using parameters estimated using data DB (respectively DA). Once again if the model were perfect, the points would be close to the 45° line.

As suggested by the values of the log-likelihood $\tilde{\ell}$ in Tables 2 and 4, the figures for models using first-guess covariates indicate weaker goodness of fit and weaker cross-validation than Figures 1-24 for models with observed wind speed covariates. Both goodness-of-fit and cross-validation appear to improve somewhat for lower pressure levels; Figures 17B-24B. This suggests that models using first-guess covariates have somewhat better predictive and descriptive value at 250mb levels. However, they appear to be not as good as models using observed wind speed as covariates.

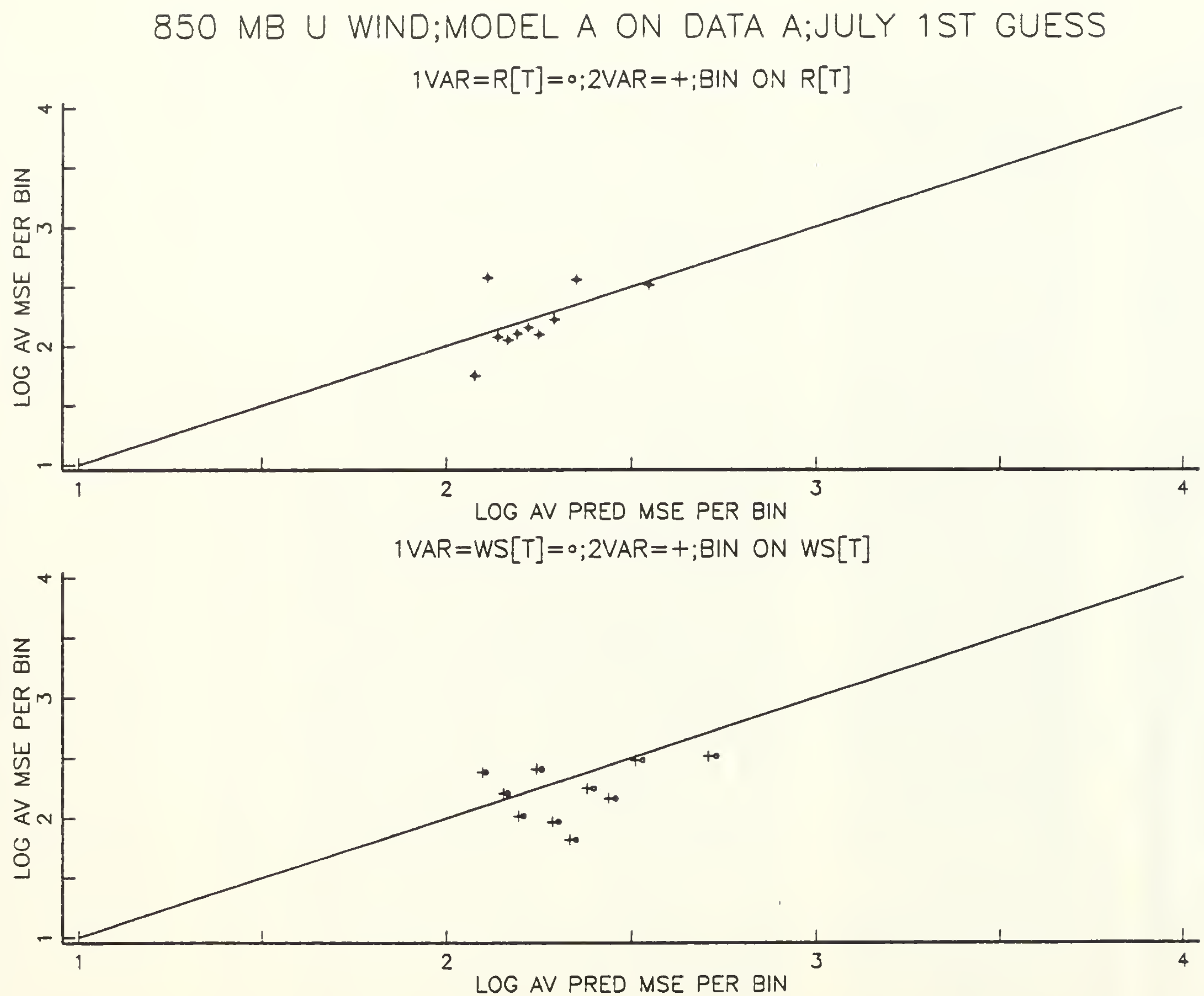

850 MB U WIND;MODEL A ON DATA A;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 1B

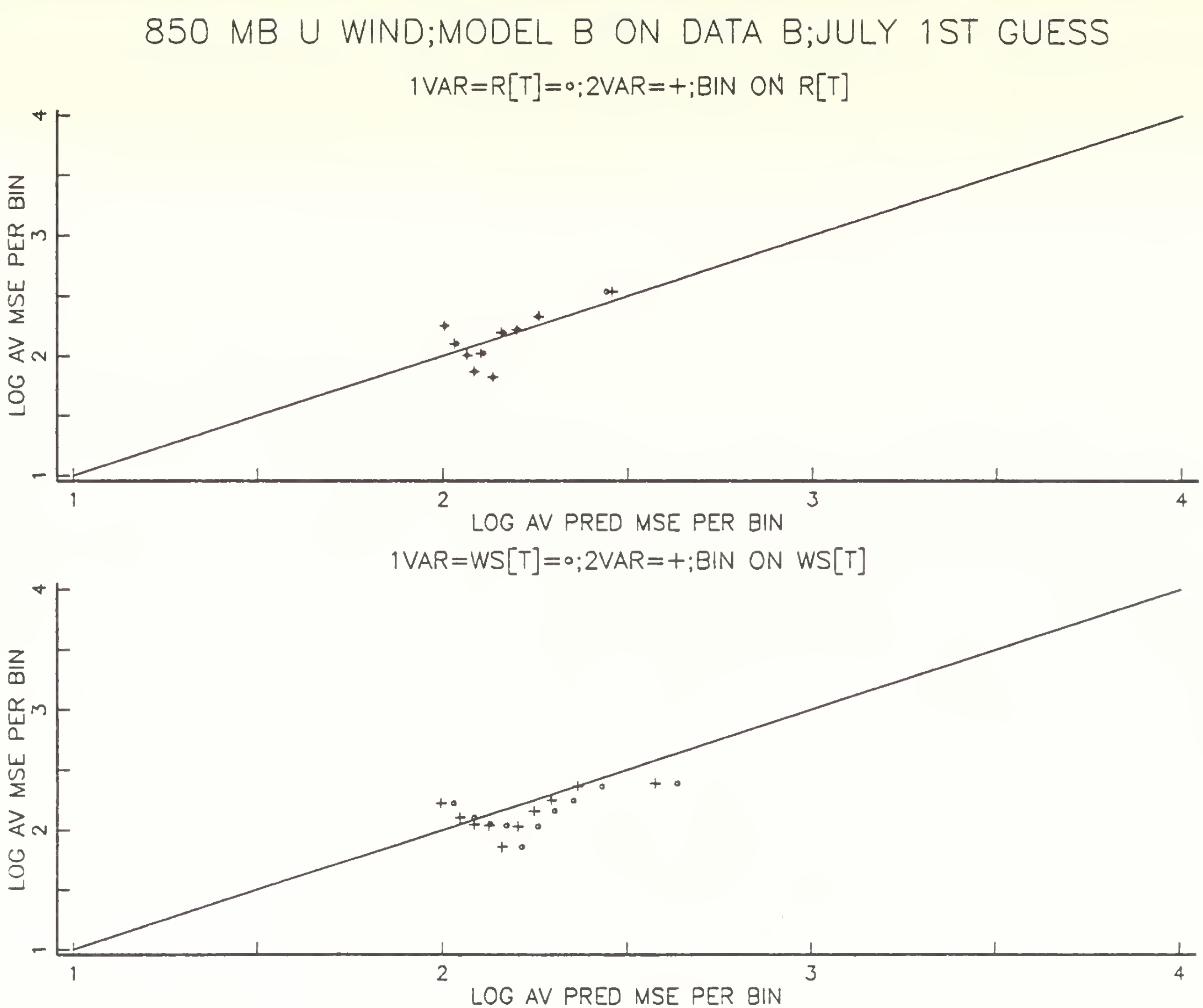

850 MB U WIND;MODEL B ON DATA B;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 2B

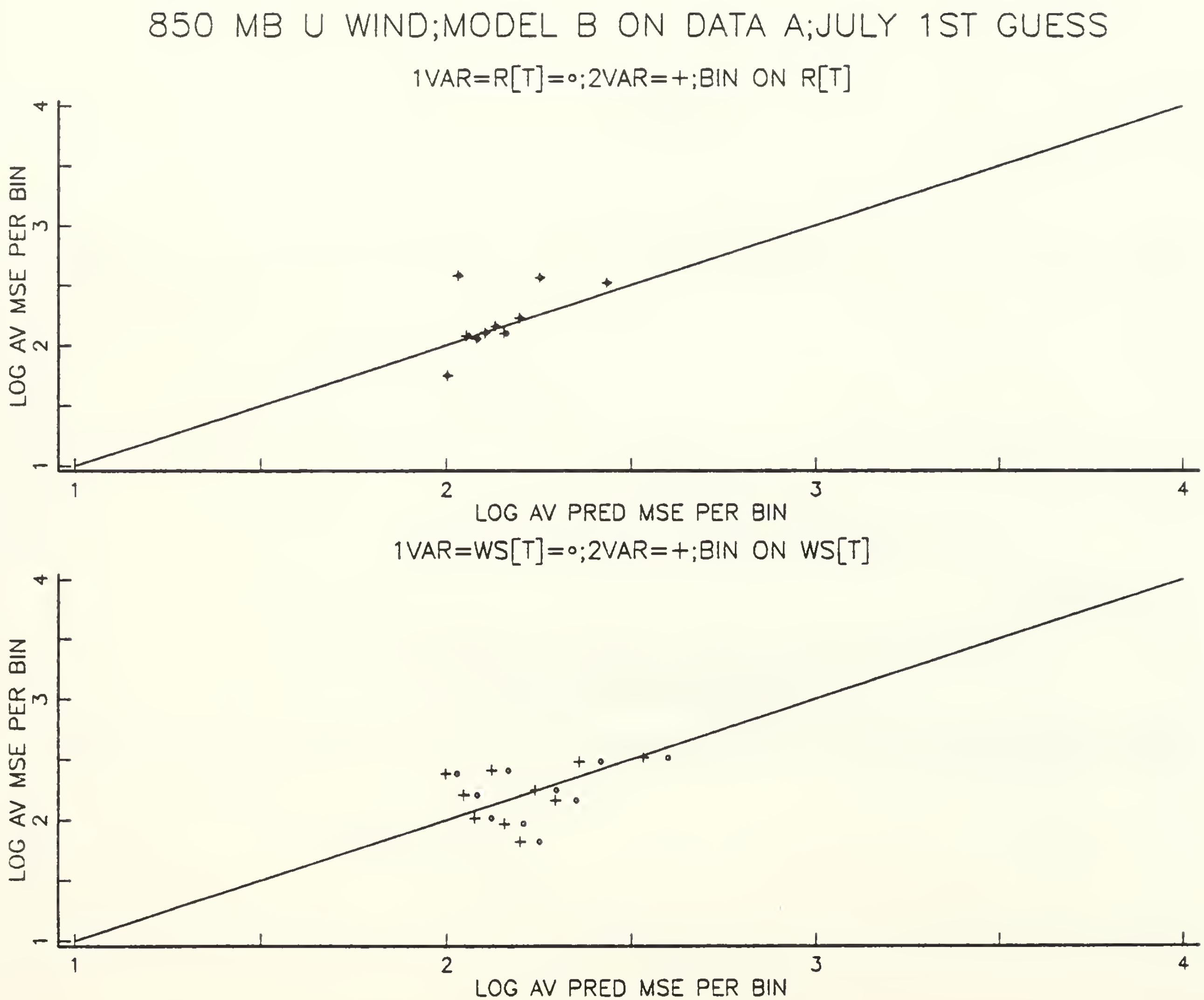

850 MB U WIND;MODEL B ON DATA A;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 3B

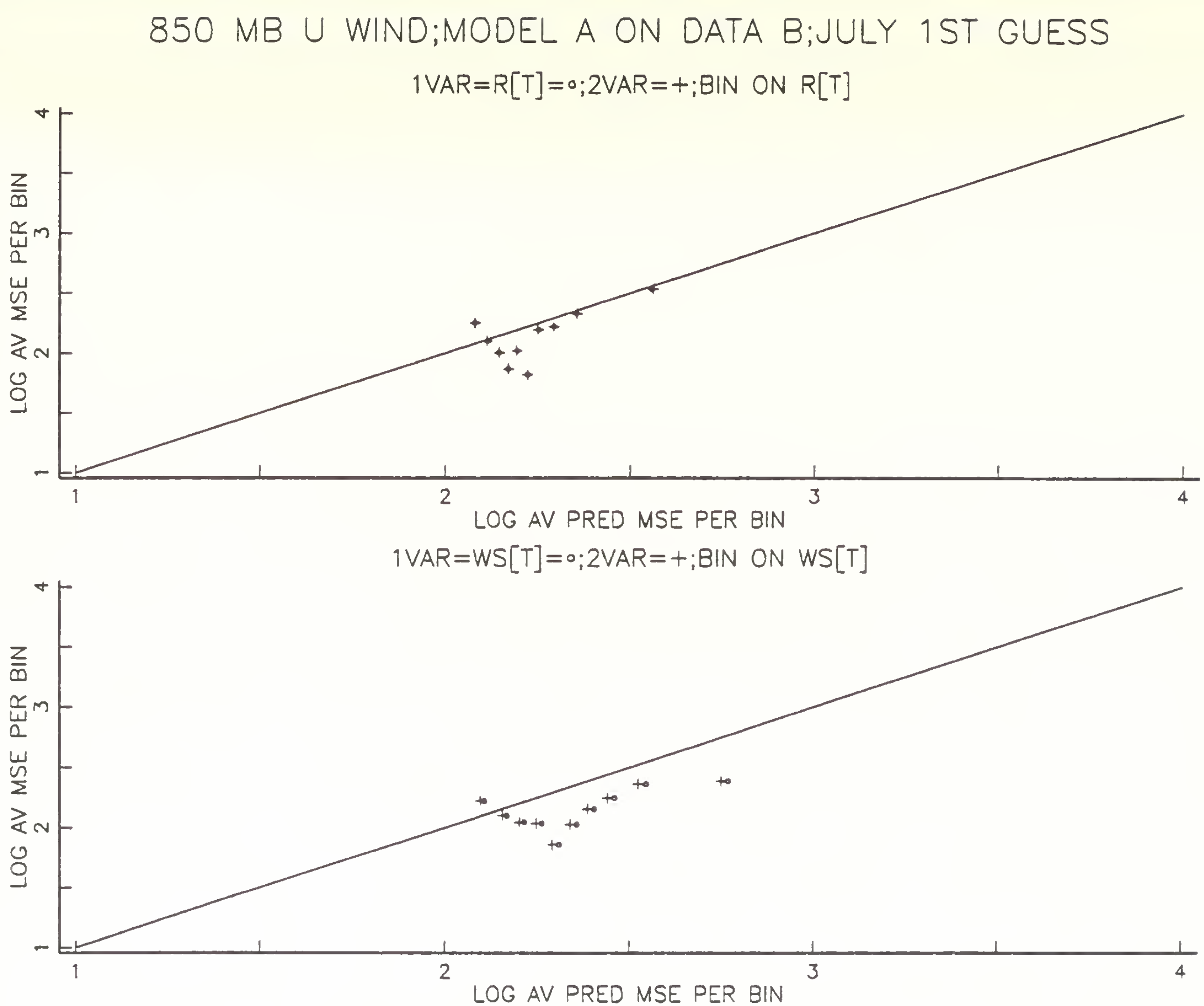

850 MB U WIND;MODEL A ON DATA B;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 4B

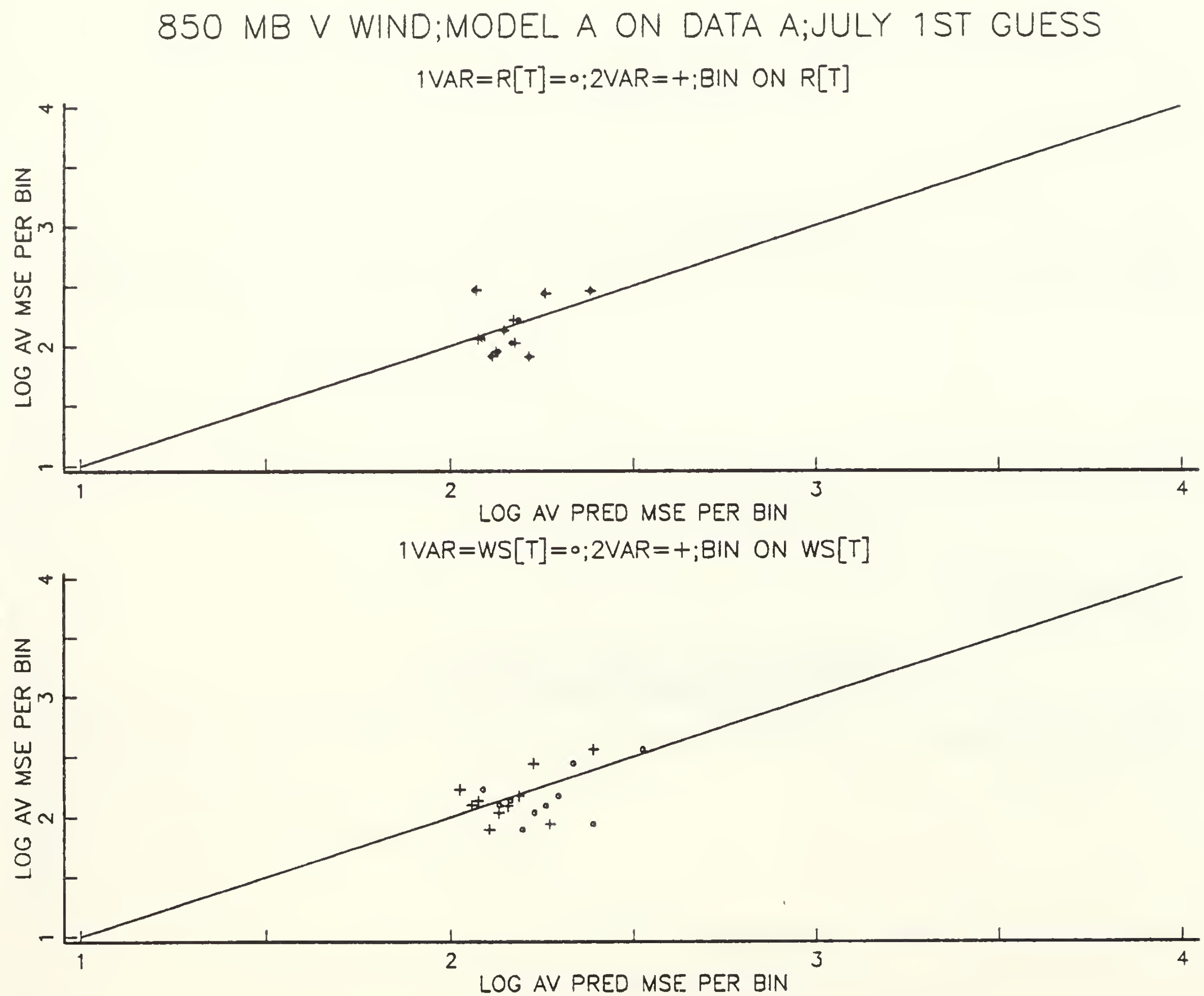

850 MB V WIND;MODEL A ON DATA A;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1
2
3
4

Figure 5B

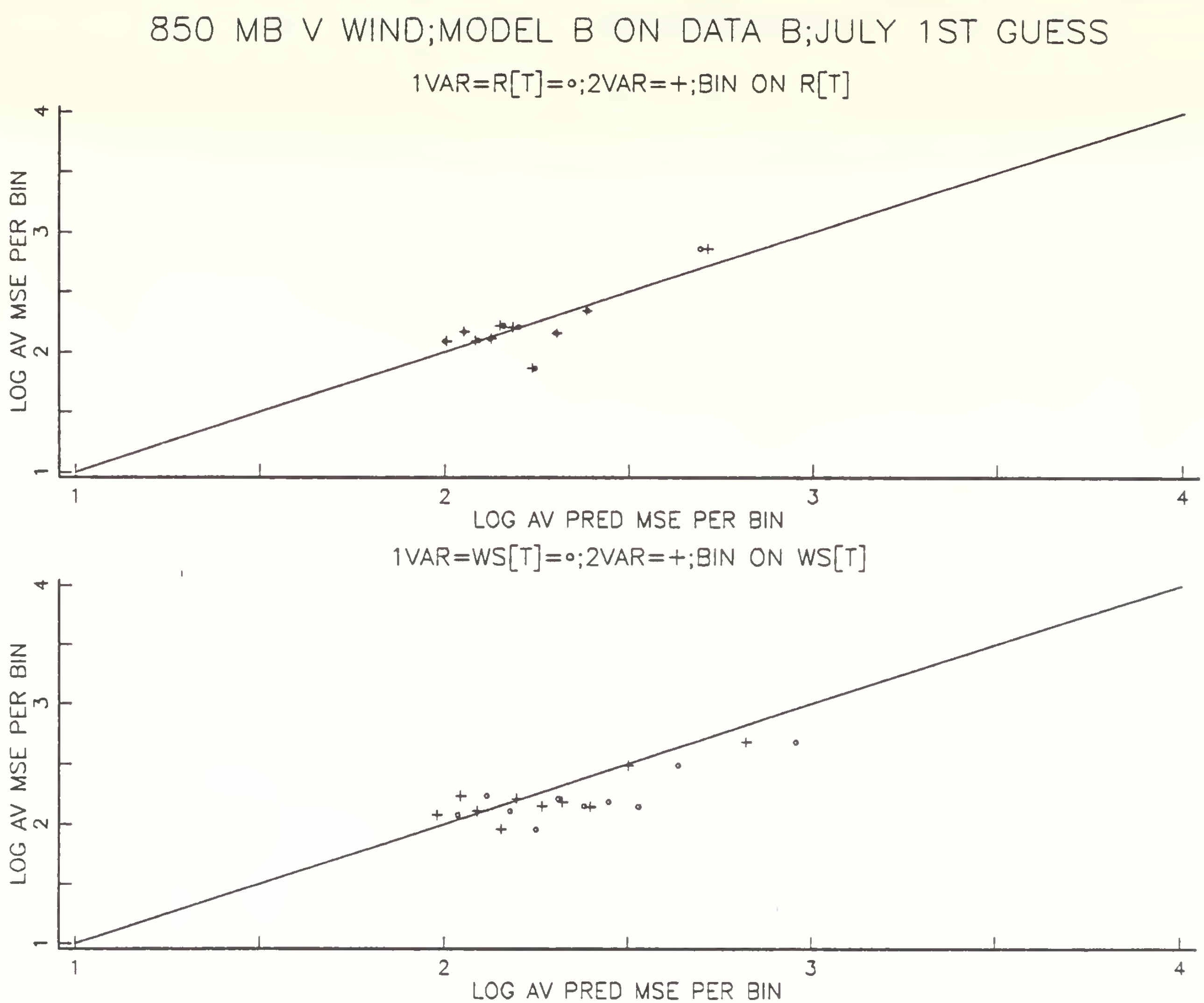
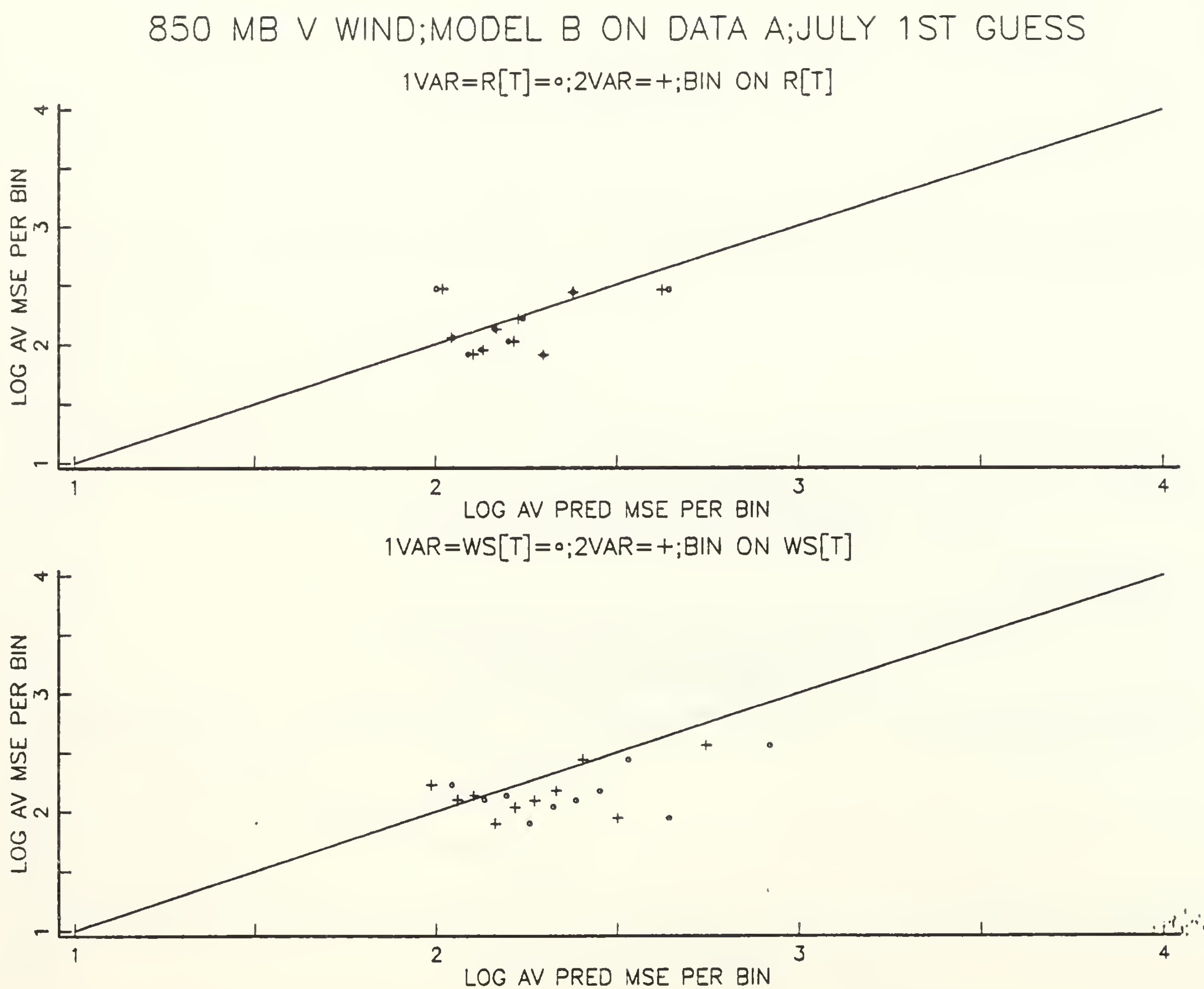
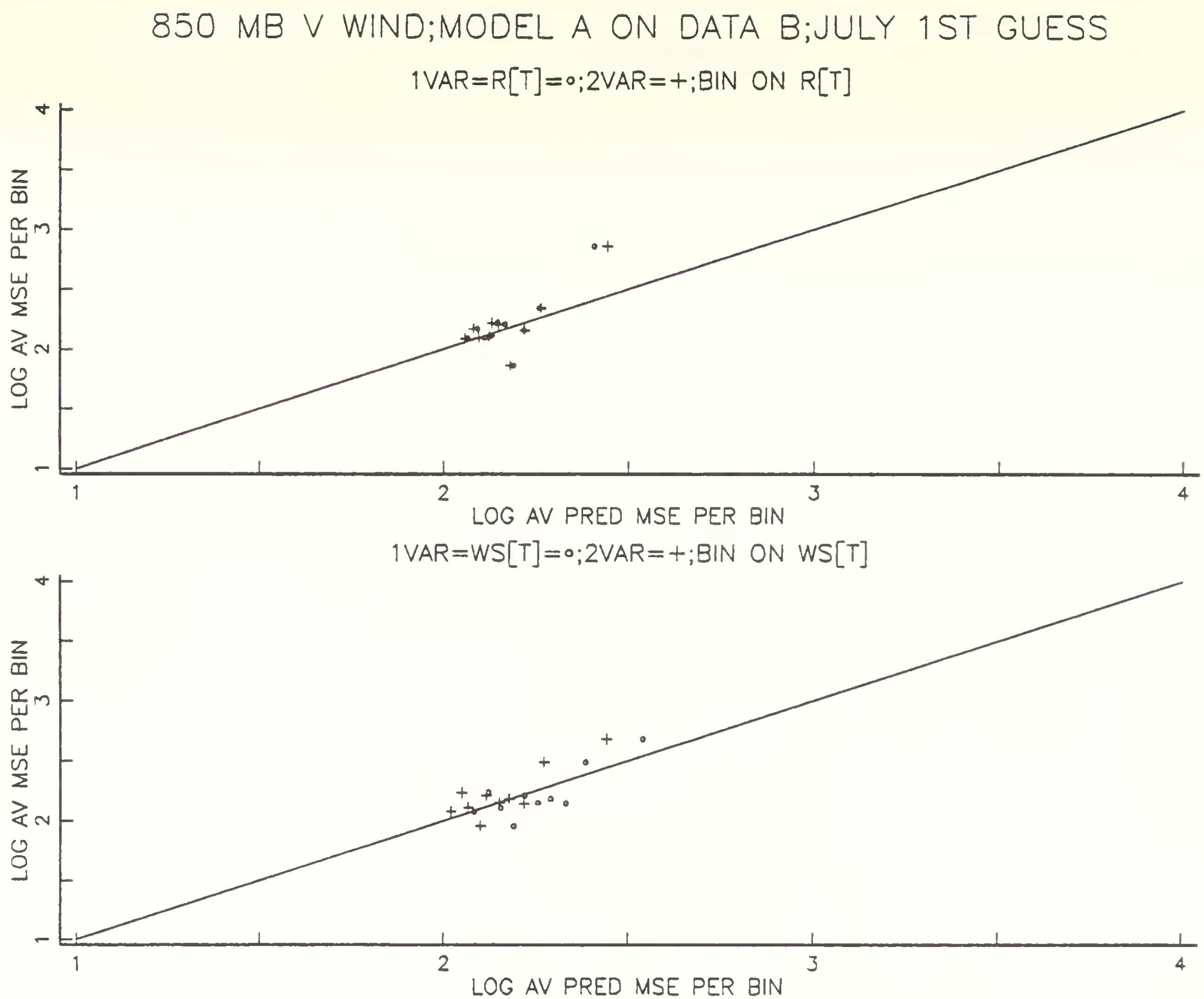

850 MB V WIND;MODEL B ON DATA B;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 6B



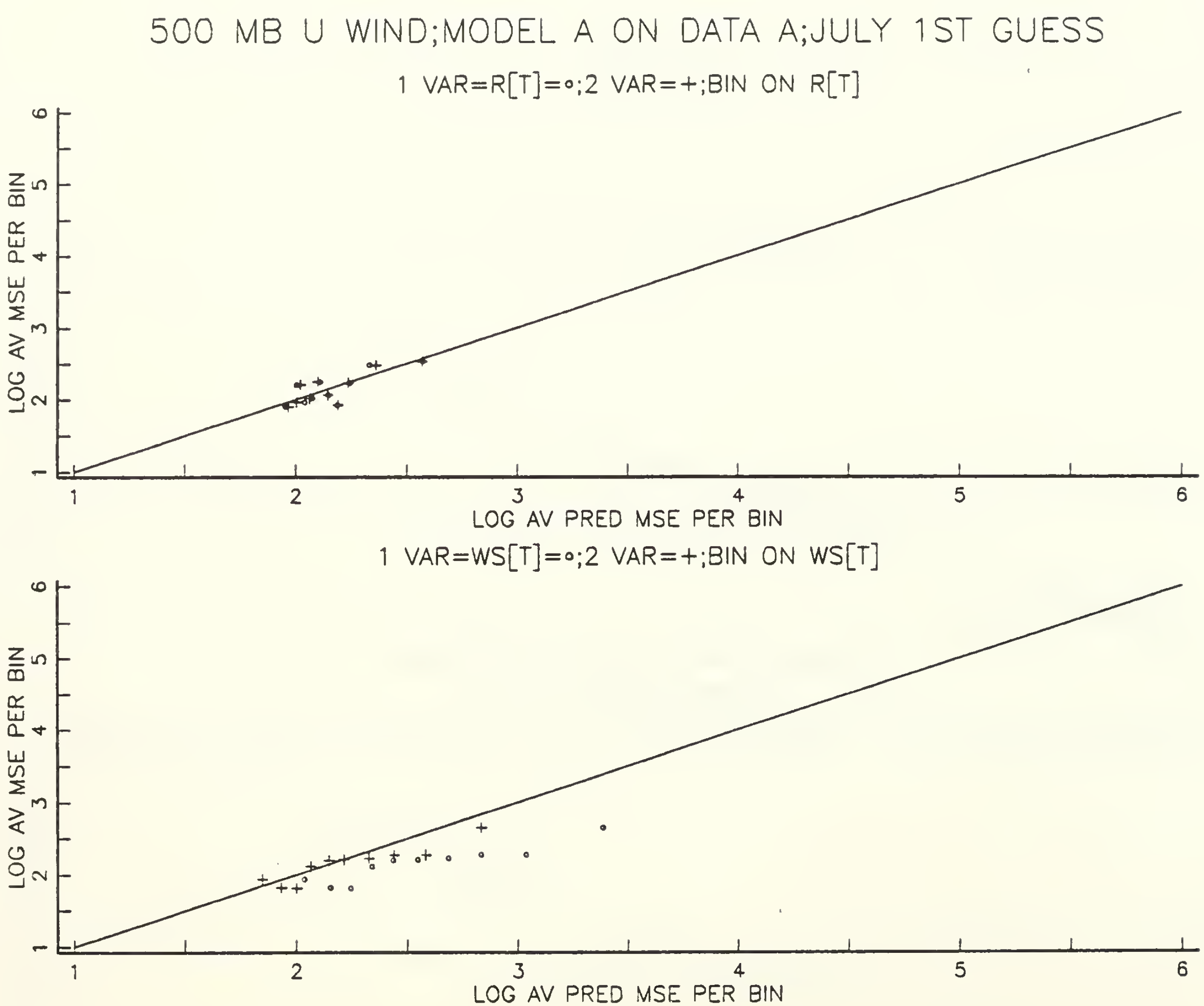
850 MB V WIND;MODEL B ON DATA A;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 7B



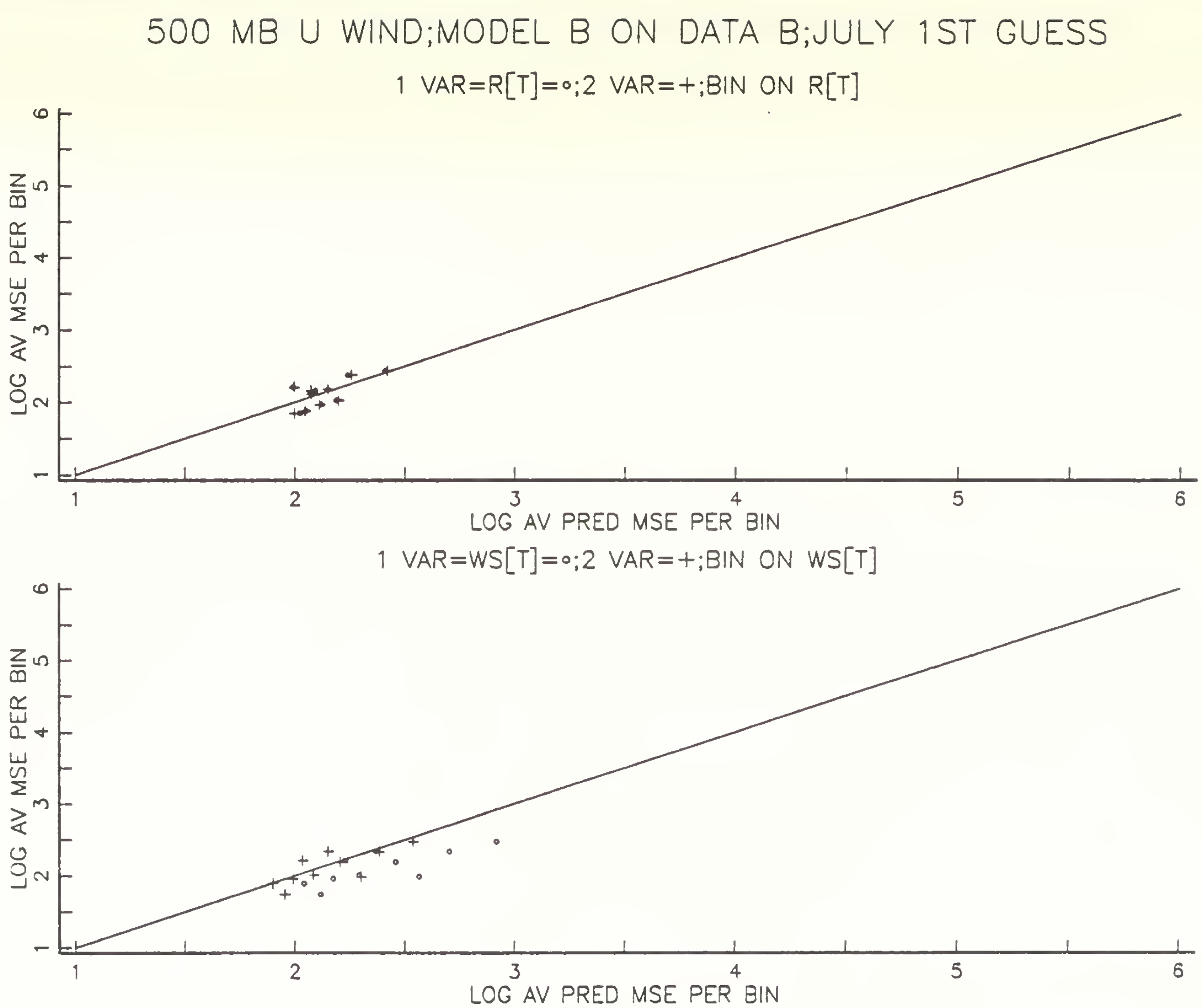
850 MB V WIND;MODEL A ON DATA B;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 8B



500 MB U WIND;MODEL A ON DATA A;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 9B



500 MB U WIND;MODEL B ON DATA B;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 10B

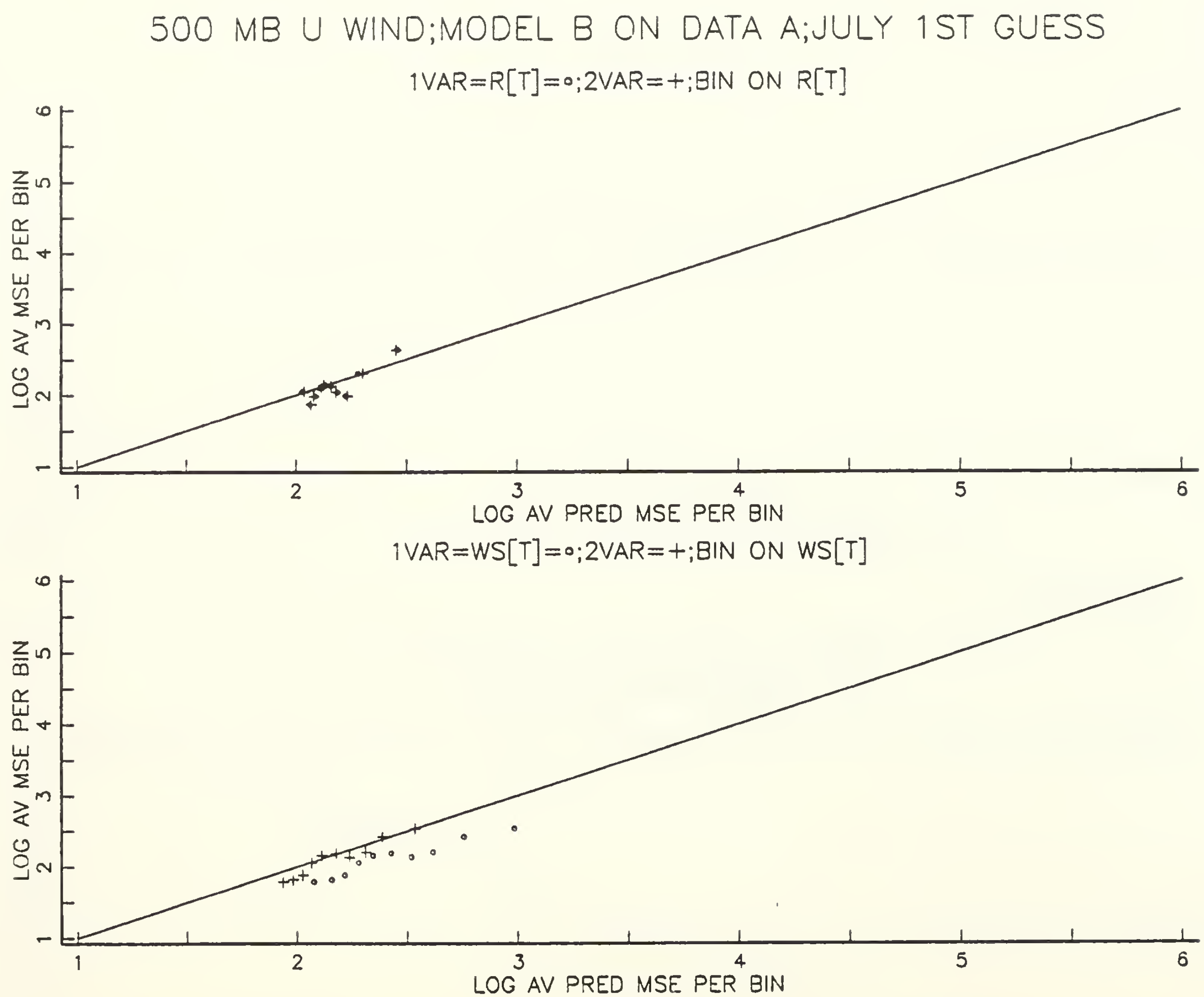

500 MB U WIND;MODEL B ON DATA A;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 11B

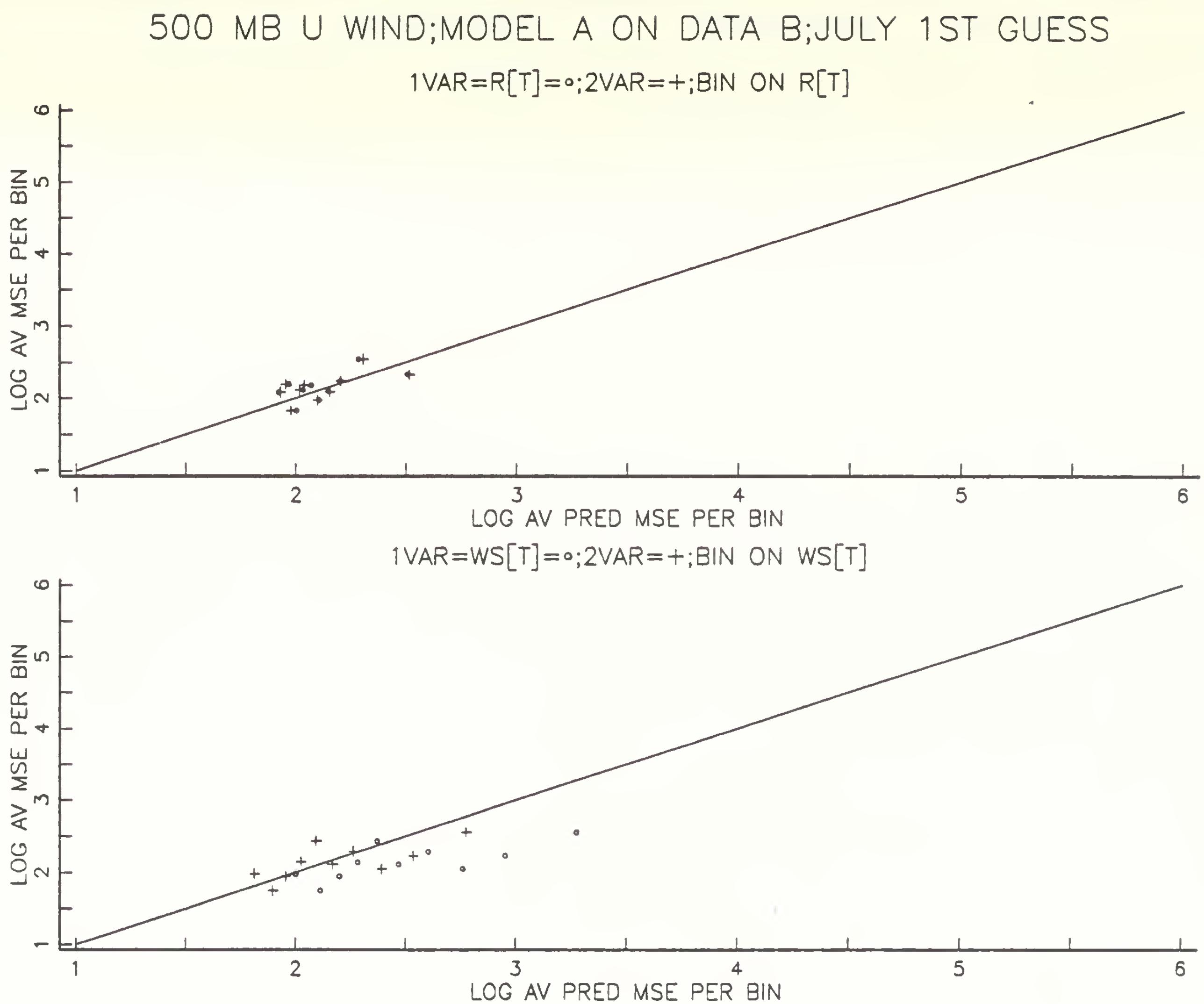

500 MB U WIND;MODEL A ON DATA B;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 12B

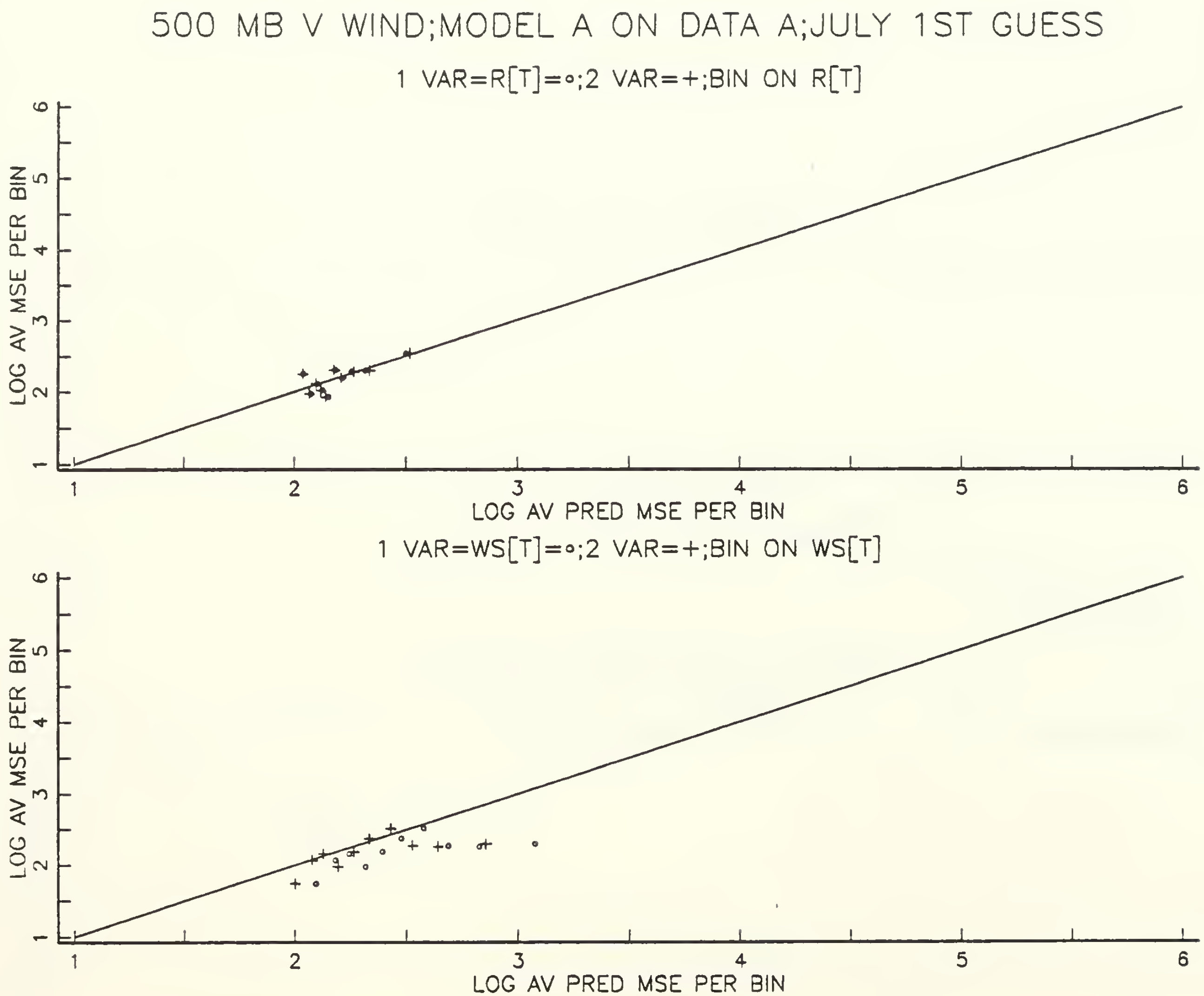

500 MB V WIND;MODEL A ON DATA A;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 13B

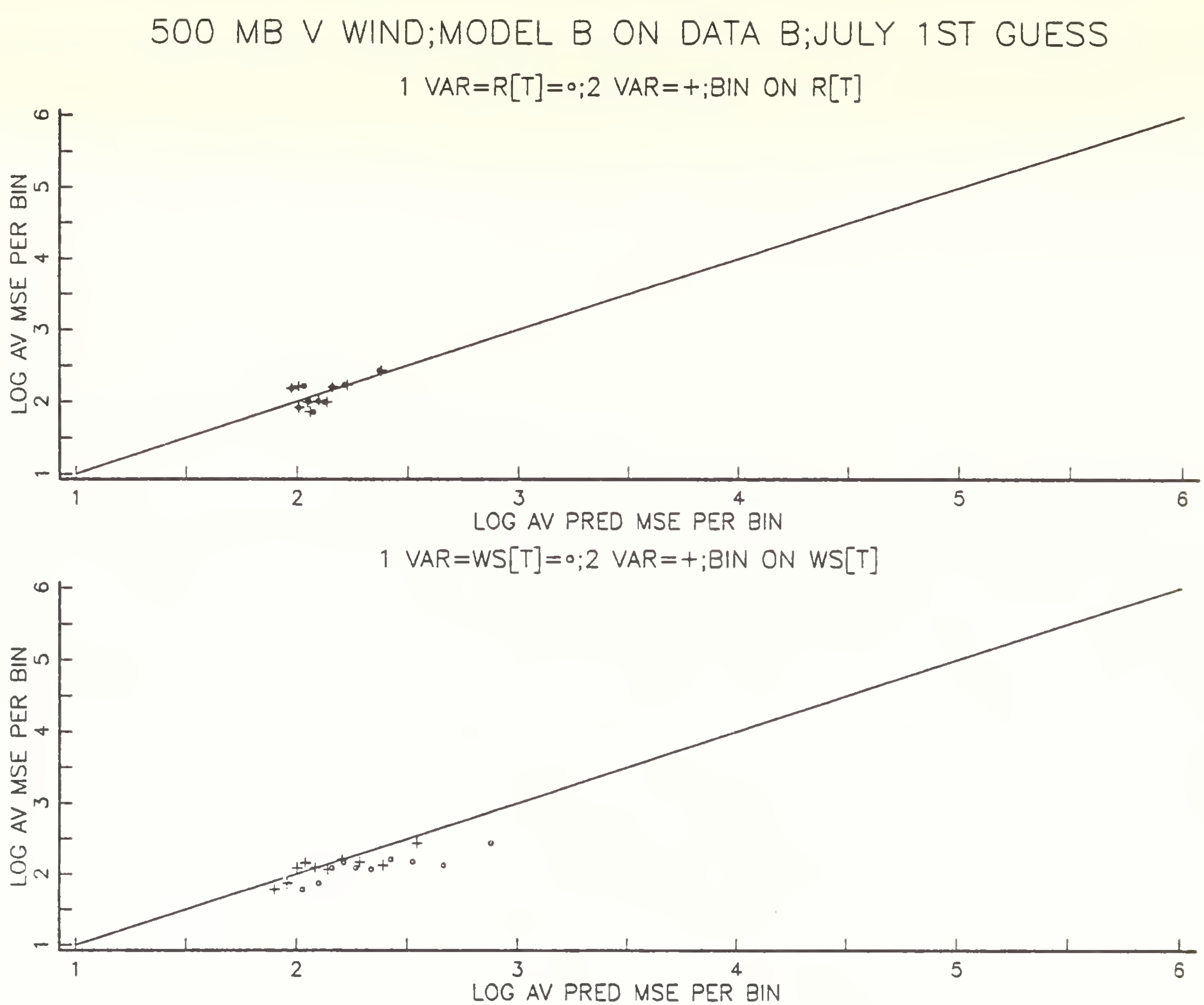

500 MB V WIND;MODEL B ON DATA B;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 14B

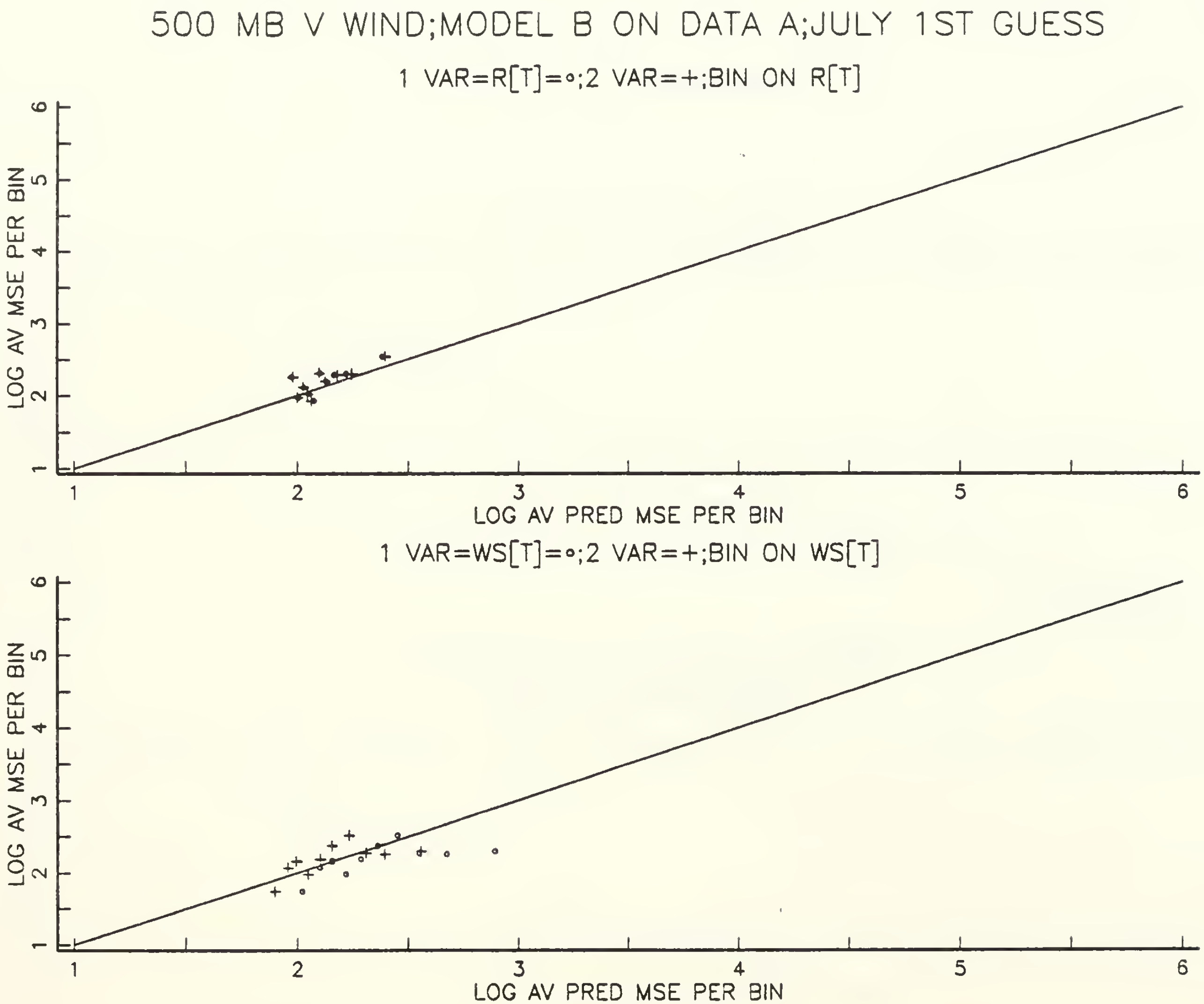

500 MB V WIND;MODEL B ON DATA A;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1
2
3
4
5
6

Figure 15B

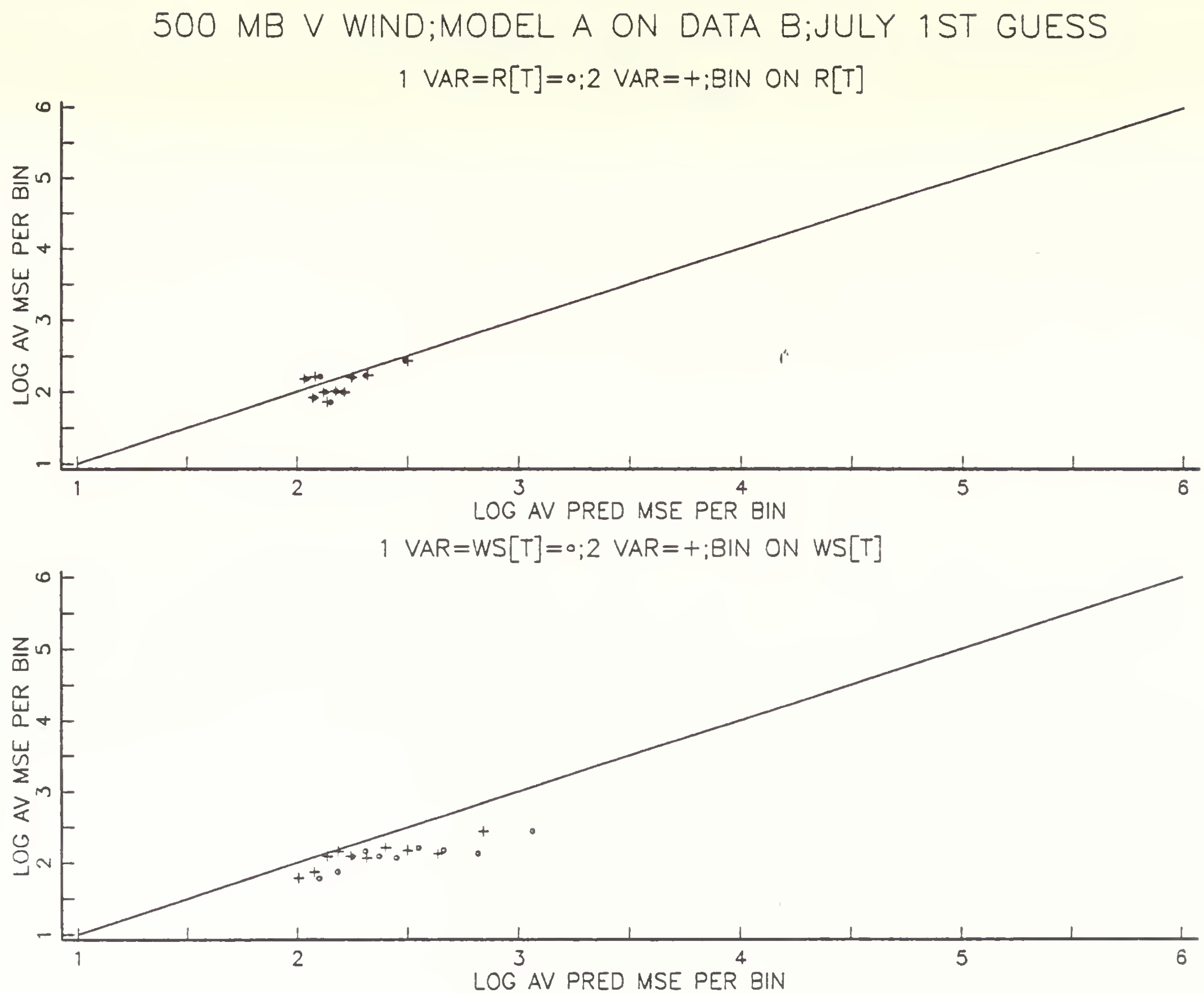

500 MB V WIND;MODEL A ON DATA B;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 16B

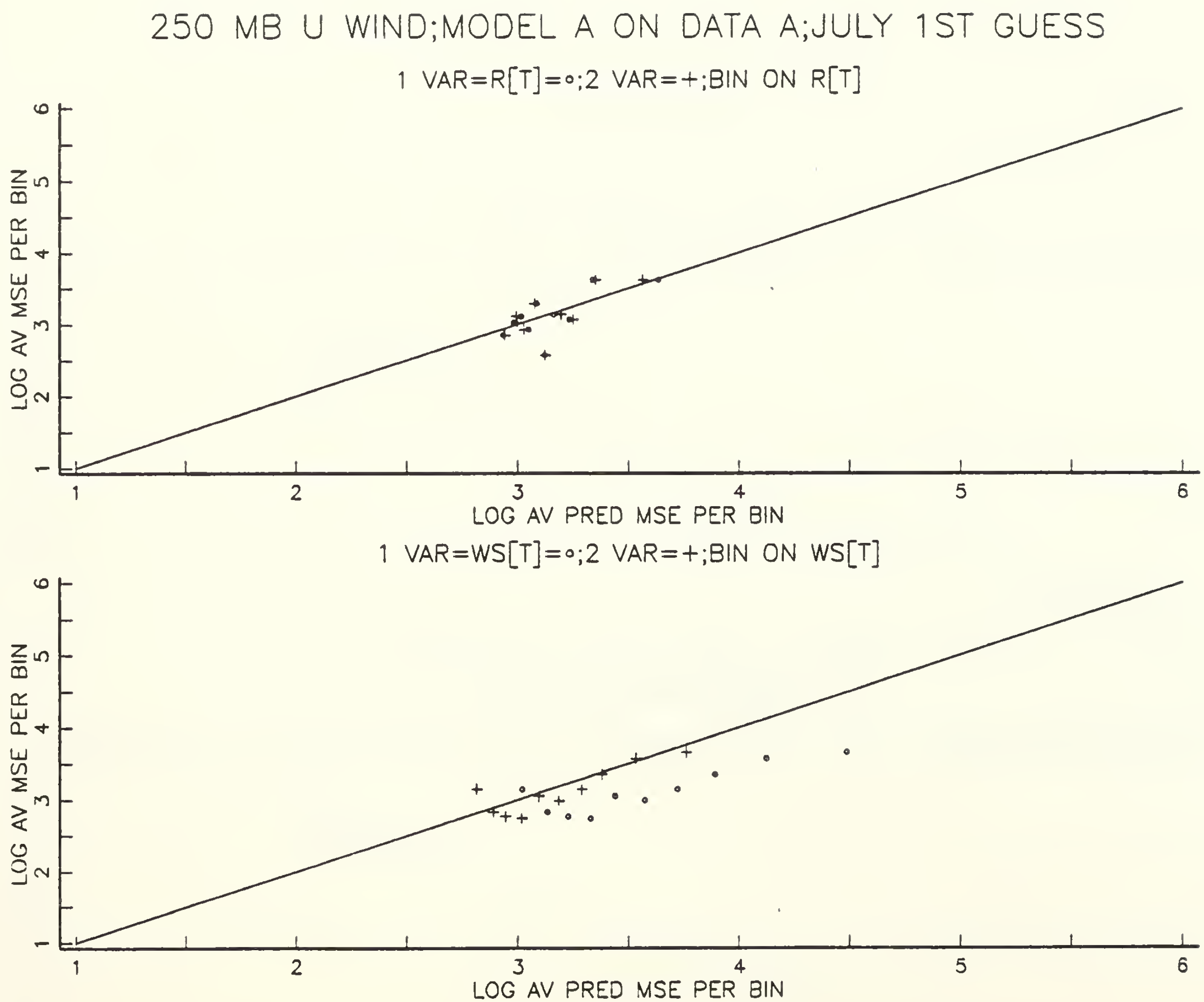

250 MB U WIND;MODEL A ON DATA A;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 17B

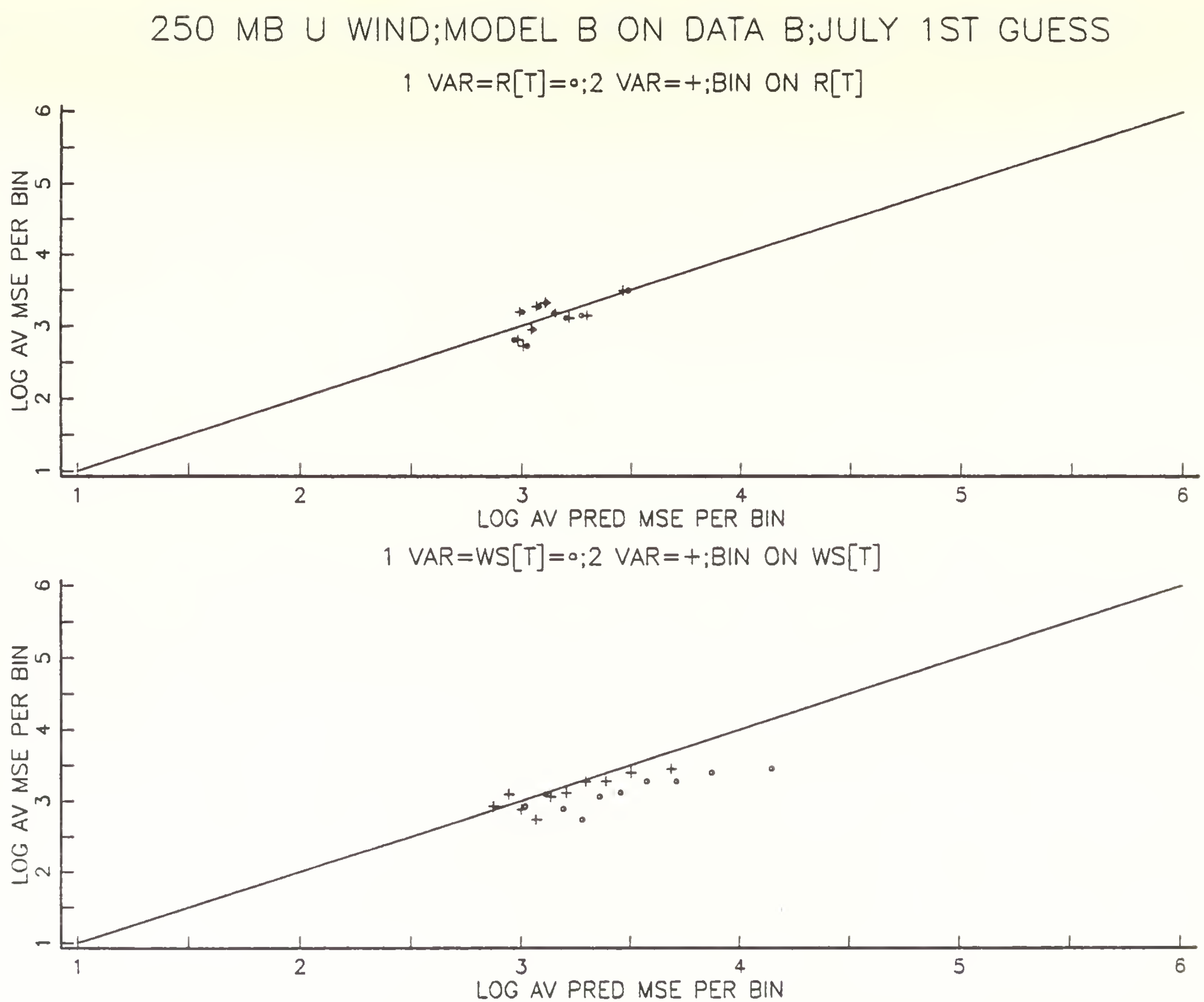

250 MB U WIND;MODEL B ON DATA B;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 18B

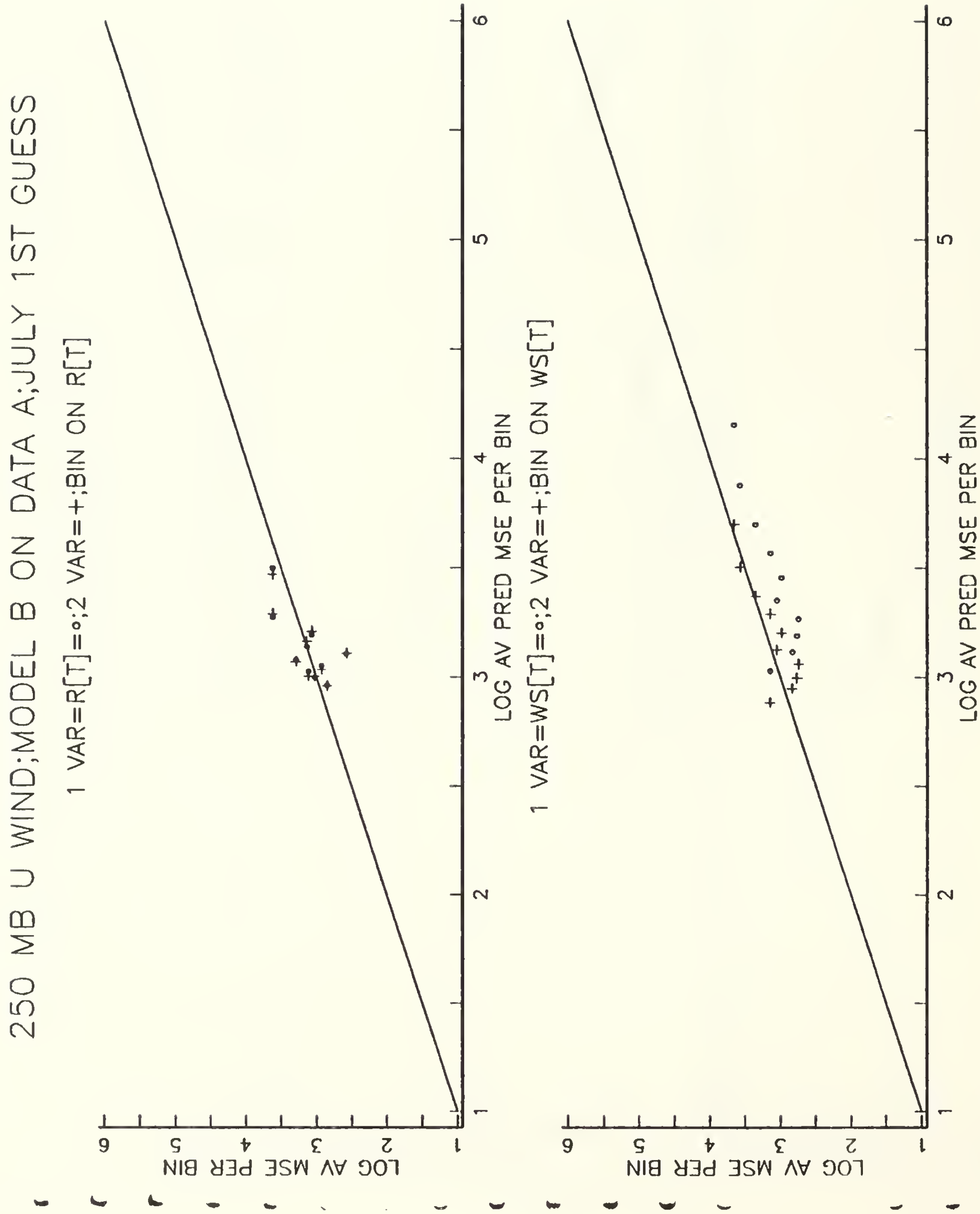

250 MB U WIND;MODEL B ON DATA A;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 19B

250 MB U WIND;MODEL A ON DATA B;JULY 1ST GUESS

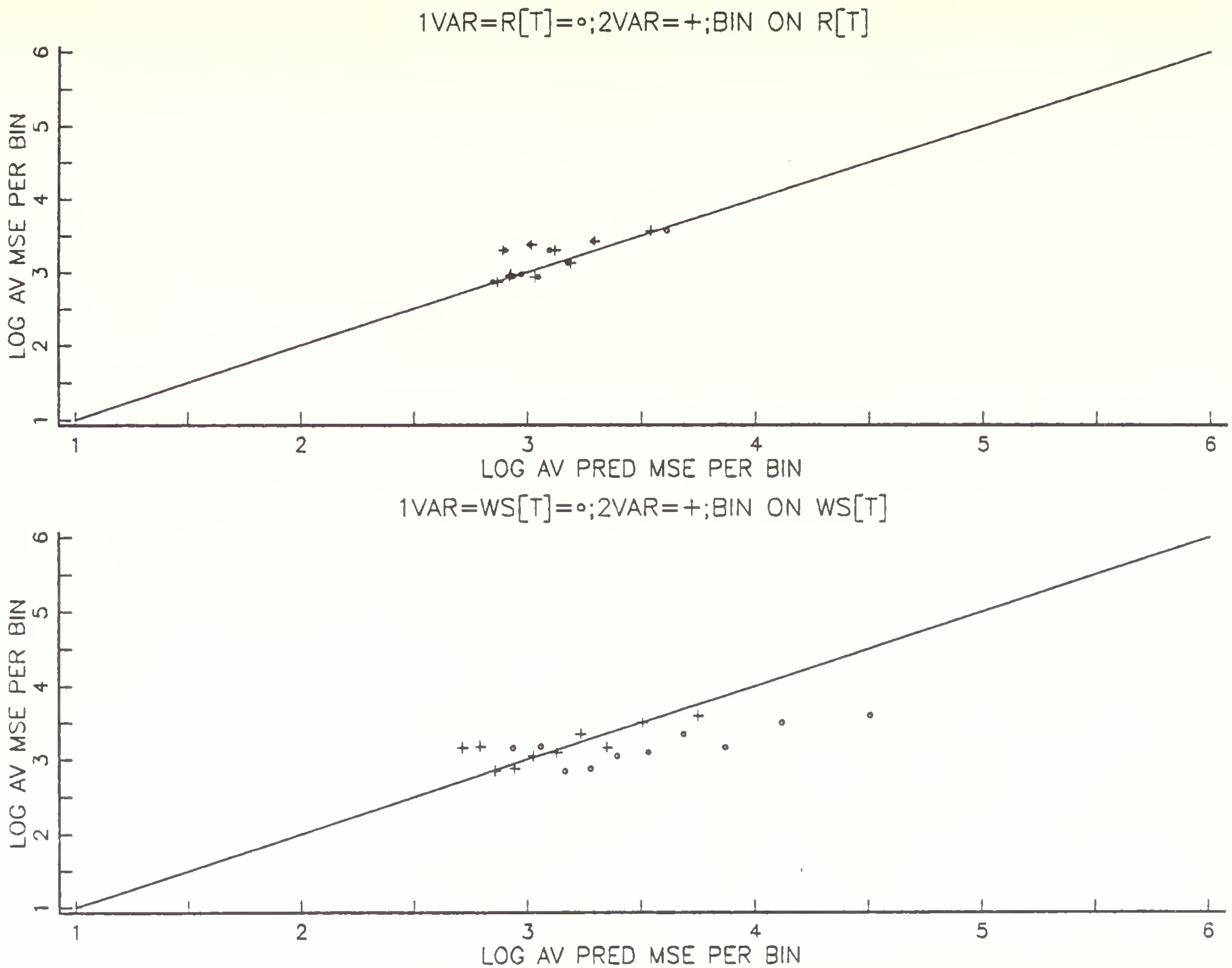


Figure 20B

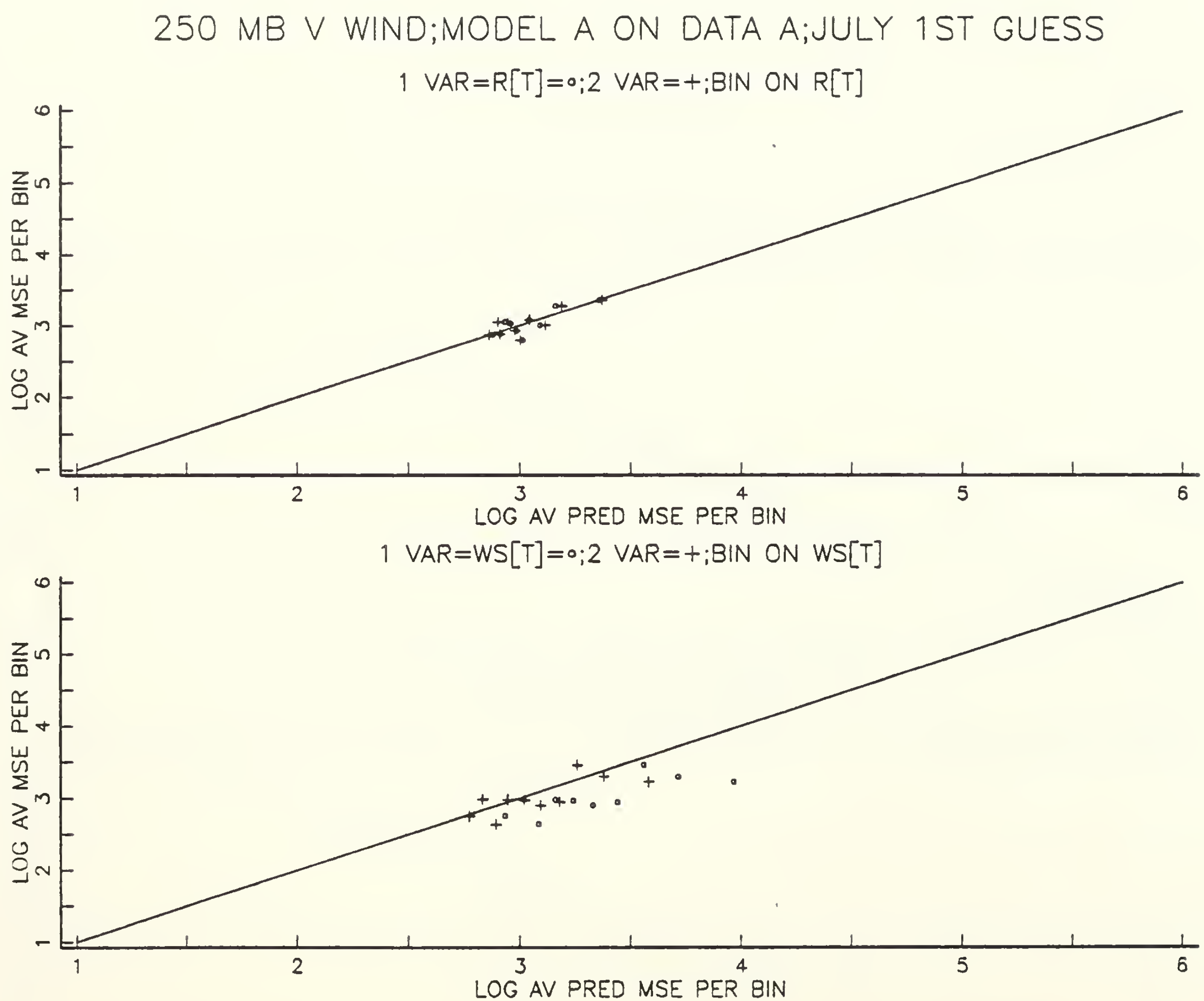

250 MB V WIND;MODEL A ON DATA A;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 2 3 4 5 6

Figure 21B

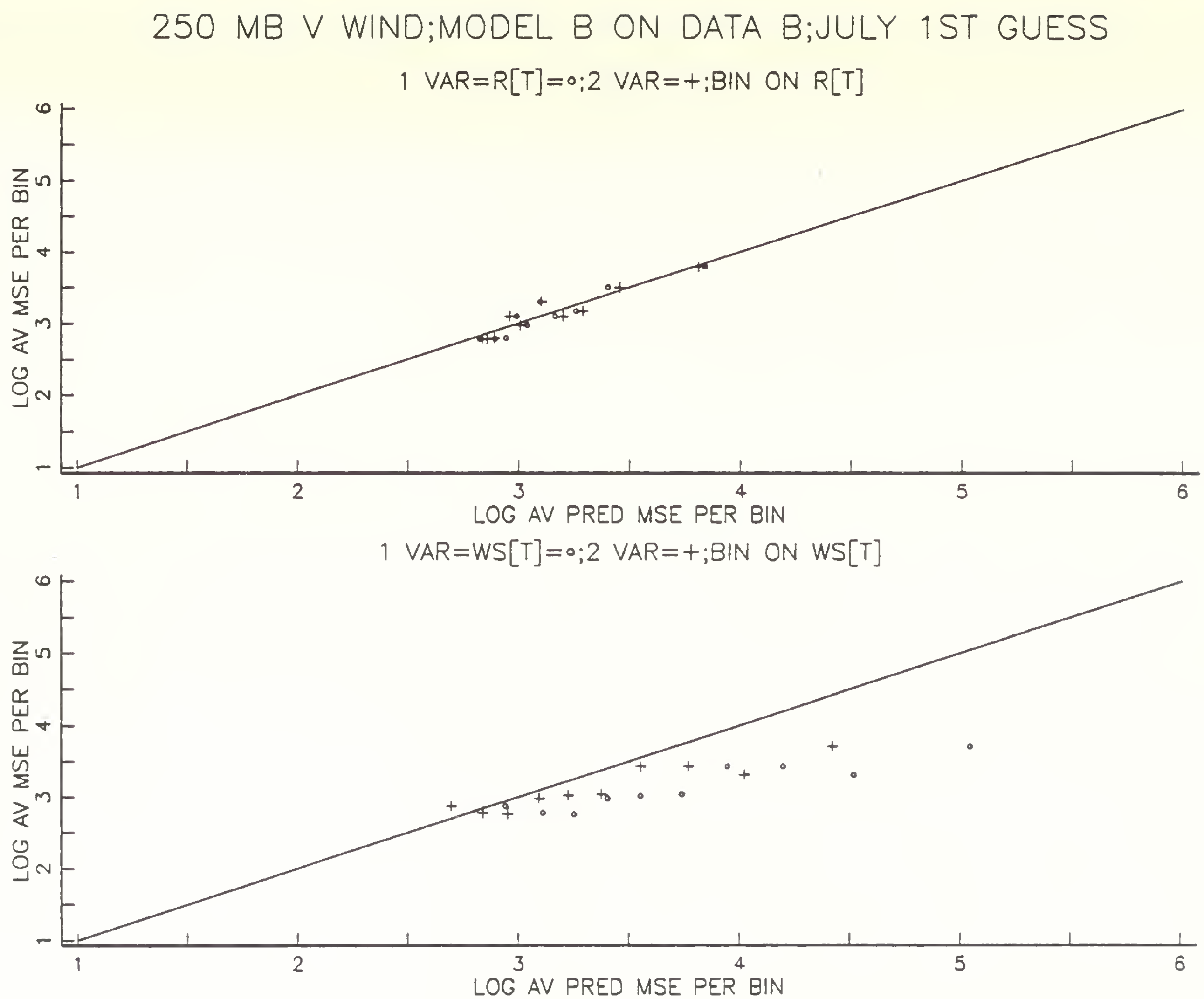

250 MB V WIND;MODEL B ON DATA B;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 22B

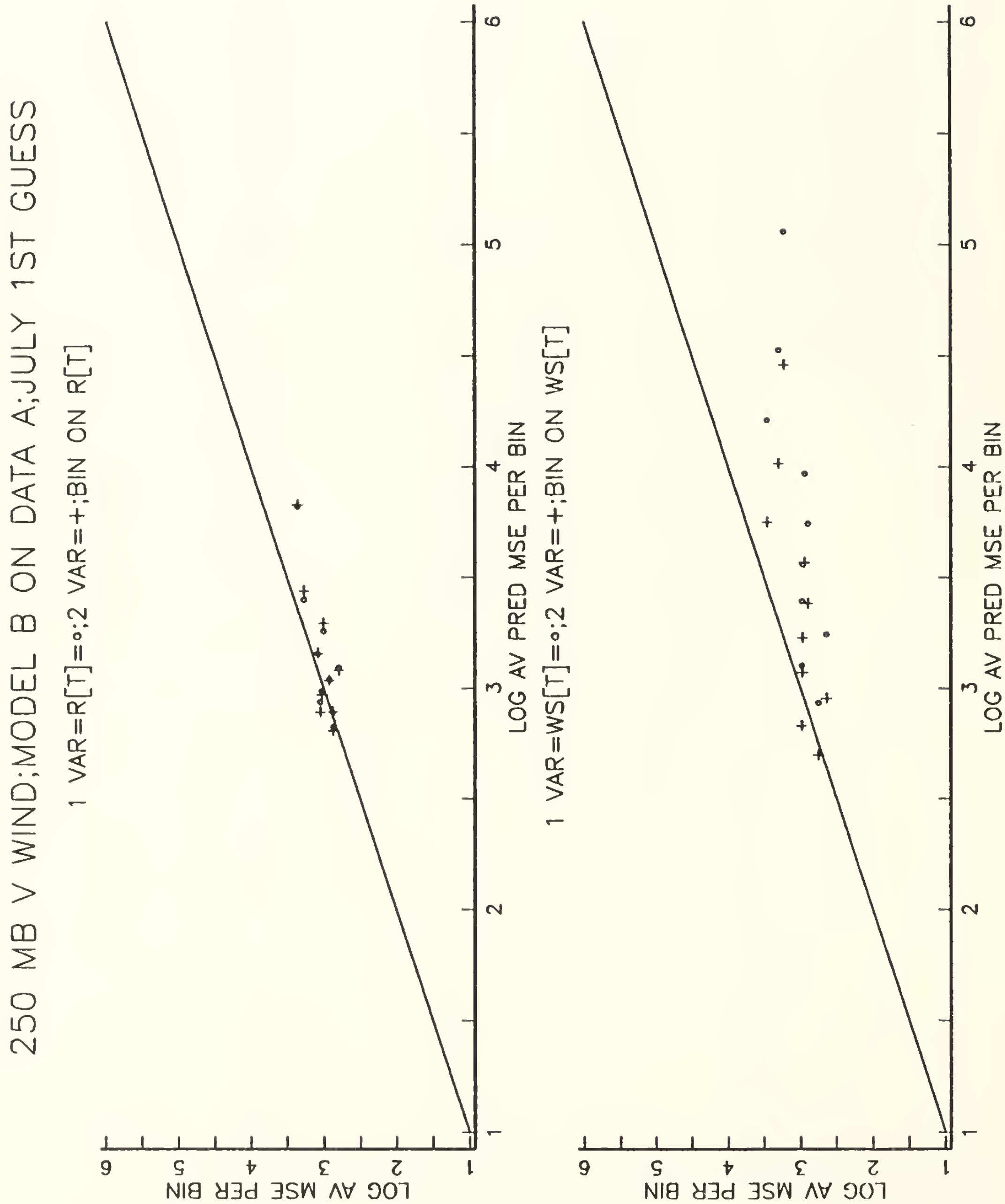

250 MB V WIND;MODEL B ON DATA A;JULY 1ST GUESS
1 VAR=R[T]=∘;2 VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1 VAR=WS[T]=∘;2 VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 23B

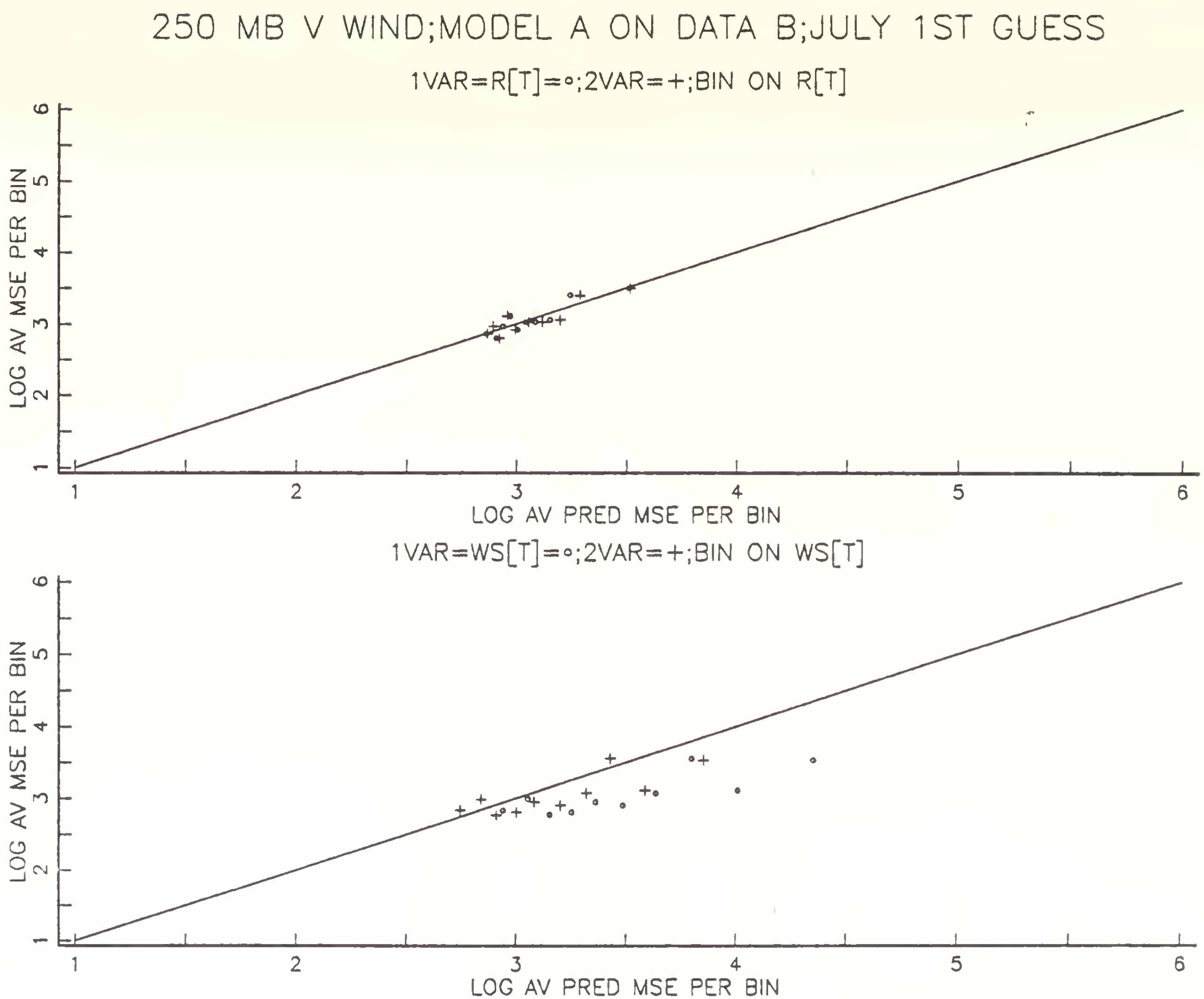

250 MB V WIND;MODEL A ON DATA B;JULY 1ST GUESS
1VAR=R[T]=∘;2VAR=+;BIN ON R[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN
1VAR=WS[T]=∘;2VAR=+;BIN ON WS[T]
LOG AV MSE PER BIN
LOG AV PRED MSE PER BIN

Figure 24B

INITIAL DISTRIBUTION LIST

1. Library (Code 52)..2
   Naval Postgraduate School
   Monterey, CA 93943-5000

2. Defense Technical Information Center...2
   Cameron Station
   Alexandria, VA 22314

3. Office of Research Administration (Code 81)..1
   Naval Postgraduate School
   Monterey, CA 93943-5000

4. Prof. Peter Purdue...1
   Code OR-Pd
   Naval Postgraduate School
   Monterey, CA 93943-5000

5. Department of Operations Research (Code 55)..1
   Naval Postgraduate School
   Monterey, CA 93943-5000

6. Prof. Donald Gaver, Code OR-Gv...5
   Naval Postgraduate School
   Monterey, CA 93943-5000

7. Prof. Patricia Jacobs..5
   Code OR/Jc
   Naval Postgraduate School
   Monterey, CA 93943-5000

8. Center for Naval Analyses..1
   4401 Ford Avenue
   Alexandria, VA 22302-0268

9. Dr. David Brillinger...1
   Statistics Department
   University of California
   Berkeley, CA 94720

10. Prof. Bernard Harris 1
Dept of Statistics
University of Wisconsin
610 Walnut Street
Madison, WI 53706

11. Prof. W. R. Schucany...1
Dept. of Statistics
Southern Methodist University
Dallas, TX 75222

12. Prof. D. C. Siegmund...1
Dept. of Statistics
Sequoia Hall
Stanford University
Stanford, CA 94305

13. Prof. H. Solomon 1
Department of Statistics
Sequoia Hall
Stanford University
Stanford, CA 94305

14. Dr. Ed Wegman...1
George Mason University
Fairfax, VA 22030

15. Dr. P. Welch 1
IBM Research Laboratory
Yorktown Heights, NY 10598

16. Dr. Neil Gerr 1
Office of Naval Research
Arlington, VA 22217

17. Prof. Roy Welsch...1
Sloan School
M.I.T.
Cambridge, MA 02139

18. Dr. J. Abrahams, Code 1111, Room 607 1
Mathematical Sciences Division, Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

19. Prof. J. R. Thompson..1
Dept. of Mathematical Science
Rice University
Houston, TX 77001

20. Prof. M. Leadbetter..1
Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

21. Prof. J. B. Kadane...1
Dept. of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213

22. Prof. J. Lehoczky..1
Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213

23. Prof. M. Mazumdar...1
Dept. of Industrial Engineering
University of Pittsburgh
Pittsburgh, PA 15235

24. Prof. M. Rosenblatt...1
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093

25. Prof. H. Chernoff..1
Department of Statistics
Harvard University
1 Oxford Street
Cambridge, MA 02138

26. Prof. Frank Samaniego...1
Statistics Department
University of California
Davis, CA 95616

27. Dr. James McKenna 1
Bell Communications Research
445 South Street
Morristown, NJ 07960-1910

28. Prof. Brad Carlin 1
School of Public Health
University of Minnesota
Mayo Bldg. A460
Minneapolis, MN 55455

29. Prof. W. Stuetzle 1
Department of Statistics
University of Washington
Seattle, WA 98195

30. Prof. F. W. Mosteller 1
Department of Statistics
Harvard University
1 Oxford St.
Cambridge, MA 02138

31. Dr. D. C. Hoaglin 1
Department of Statistics
Harvard University
1 Oxford Street
Cambridge, MA 02138

32. Prof. N. D. Singpurwalla 1
George Washington University
Washington, DC 20052

33. Prof. George S. Fishman 1
Curr. in OR & Systems Analysis
University of North Carolina
Chapel Hill, NC 20742

34. Prof. Bradley Efron 1
Statistics Dept.
Sequoia Hall
Stanford University
Stanford, CA 94305

35. Prof. Carl N. Morris 1
Statistics Department
Harvard University
1 Oxford St.
Cambridge, MA 02138

36. Dr. John E. Rolph 1
RAND Corporation
1700 Main St.
Santa Monica, CA 90406

37. Dr. Edward G. Coffman, Jr 1
AT&T Bell Telephone Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

38. Operations Research Center, Rm E40-164 1
Massachusetts Institute of Technology
Attn: R. C. Larson and J. F. Shapiro
Cambridge, MA 02139

39. Arthur P. Hurter, Jr. 1
Professor and Chairman
Dept. of Industrial Engineering and Management Sciences
Northwestern University
Evanston, IL 60201-9990

40. Institute for Defense Analysis 1
1800 North Beauregard
Alexandria, VA 22311

41. Prof. J. W. Tukey 1
Statistics Dept., Fine Hall
Princeton University
Princeton, NJ 08540

42. Dr. Colin Mallows 1
AT&T Bell Telephone Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

43. Dr. Jon Kettenring..1
Bellcore
445 South Street
Morris Township, NJ 07962-1910

44. Prof. David L. Wallace 1
Statistics Dept., University of Chicago
5734 S. University Ave.
Chicago, IL 60637

45. Dr. S. R. Dalal..1
Bellcore
445 South Street
Morristown, NJ 07962-1910

46. Dr. D. F. Daley 1
Statistic Dept. (I.A.S.)
Australian National University
Canberra, A.C.T. 2606
AUSTRALIA

47. Dr. Guy Fayolle...1
I.N.R.I.A.
Dom de Voluceau-Rocquencourt
78150 Le Chesnay Cedex
FRANCE

48. Professor H. G. Daellenbach..1
Department of Operations Research
University of Canterbury
Christchurch, NEW ZEALAND

49. Koh Peng Kong..1
OA Branch, DSO
Ministry of Defense
Blk 29 Middlesex Road
SINGAPORE 1024

50. Professor Sir David Cox 1
Nuffield College
Oxford, OXI INF
ENGLAND

51. Dr. A. J. Lawrence 1
Dept. of Mathematics,
University of Birmingham
P. O. Box 363
Birmingham B15 2TT
ENGLAND

52. Dr. F. P. Kelly 1
Statistics Laboratory
16 Mill Lane
Cambridge
ENGLAND

53. Dr. John Copas 1
Dept. of Mathematics,
University of Birmingham
P. O. Box 363
Birmingham B15 2TT
ENGLAND

54. Dr. D. Vere-Jones 1
Dept. of Math, Victoria Univ. of Wellington
P. O. Box 196
Wellington
NEW ZEALAND

55. Prof. Guy Latouche 1
University Libre Bruxelles
C.P. 212, Blvd. De Triomphe
B-1050 Bruxelles
BELGIUM

56. Dr. J. Goerss 10
Naval Oceanographic and Atmospheric Laboratory
Monterey, CA 93943-5006